



**UNIVERSIDAD DE CHILE**  
**FACULTAD DE CIENCIAS FÍSICAS Y MATEMÁTICAS**  
**DEPARTAMENTO DE INGENIERÍA INDUSTRIAL**

**FISCALIZACIÓN PREVENTIVA DEL USO DE RECURSOS DE LA SUBVENCIÓN  
ESCOLAR PREFERENCIAL, EN LA SUPERINTENDENCIA DE EDUCACIÓN**

*PROYECTO DE GRADO PARA OPTAR AL GRADO DE MAGÍSTER EN  
INGENIERÍA DE NEGOCIOS CON TECNOLOGÍAS DE INFORMACIÓN*

**CRISTIAN MARCELO SUÁREZ ROJAS**

PROFESOR GUÍA:  
CRISTIÁN ANDRÉS JULIO AMDAN

MIEMBROS DE LA COMISIÓN:  
JAIME CONTESSE MARROQUÍN  
SEBASTIÁN RÍOS PÉREZ

SANTIAGO DE CHILE

2018

## RESUMEN EJECUTIVO

Producto de la Ley Sistema Nacional de Aseguramiento de la Calidad de la Educación Parvularia, Básica y Media y su Fiscalización, N° 20.529, 2011; se crea la Superintendencia de Educación la cual, mediante la fiscalización, la rendición de cuentas y la gestión de denuncias ciudadanas garantiza el cumplimiento de las condiciones necesarias para que exista calidad y equidad en la educación parvularia y escolar.

En la actualidad, de acuerdo al artículo 48 de la Ley N° 20.529, la SUPEREDUC debe fiscalizar el cumplimiento normativo educacional de más de 1.800 obligaciones y velar por la legalidad del uso de más de 6 billones de pesos anuales (3% presupuesto total del país) que reciben los sostenedores de establecimientos educacionales como aportes estatales. Para lo anterior, cuenta con una capacidad de fiscalización limitada, que promedia las 21.400 visitas anuales y que obliga a la Organización a priorizar las materias a incluir en cada programa, así como los sostenedores y/o escuelas a fiscalizar.

Ante la problemática descrita, y tras instanciar la metodología de ingeniería de negocios se identificó que la SUPEREDUC ante el objetivo de *mejorar el uso de los recursos asignados para fines educativos*, sus procesos de fiscalización se activaban en fecha posterior a la declaración anual de gastos, y durante el año no se tomaban medidas preventivas y/o de mitigación de riesgos derivados de usos indebidos de recursos. Con esta forma de operar, la SUPEREDUC anualmente detecta miles de millones de pesos en usos indebidos que los sostenedores deben reintegrar al sistema escolar. Producto de lo cual, el propósito de la SUPEREDUC de ser garante de derechos educativos no se encuentra totalmente alineado con la cautela del debido uso de los fondos de las subvenciones de propósito específico, en especial con la subvención escolar preferencial (SEP) -recursos que por Ley deben destinarse a apoyar a los estudiantes más vulnerables del país y a garantizar que los proyectos educativos se concreten- y donde se concentra la mayor cantidad de incumplimientos en esta materia.

Por las razones expuestas, y sumado al principio de eficiencia del Estado, la SUPEREDUC debe aumentar la efectividad en las fiscalizaciones de uso de recursos y al mismo tiempo, ejecutar acciones preventivas y/o de mitigación para así, lograr disminuir año tras año el monto total en gastos no aceptados.

## **AGRADECIMIENTOS**

En primer lugar, agradezco a Nuestro Padre Celestial por darme la oportunidad de vivir esta experiencia y estar conmigo en todo este arduo proceso; solo él sabe, todos los esfuerzos realizados.

A mis ancestros, a los cuales me encomiendo a diario y siempre he confiado que desde el cielo nos cuidan como familia eterna que somos.

A mis hijos Diego y Sofía, a los cuales les he quitado en parte su tiempo, lo reclaman y con justa razón, pretendo reintegrárselo lo más rápido posible.

A mi familia y seres queridos, que siempre me han apoyado en todo lo que hago y emprendo.

A Cristián, mi profesor guía, quien dirigió mi actuar con sabiduría y mucha paciencia.

A Laura y Ana María, por toda la ayuda brindada, amplio compromiso y ser la cara amable de este proceso.

## TABLA DE CONTENIDO

<b>CAPÍTULO 1: INTRODUCCIÓN Y CONTEXTO</b> .....	<b>13</b>
1.1 ANTECEDENTES DE LA INDUSTRIA.....	13
1.1.1 Factor Político .....	13
1.1.2 Factor Económico .....	13
1.1.3 Factor Socio - Cultural.....	14
1.1.4 Factor Tecnológico.....	14
1.2 DESCRIPCIÓN GENERAL DE LA EMPRESA.....	15
1.2.1 Misión.....	16
1.2.2 Clientes .....	16
1.2.3 Servicios Principales .....	16
1.2.4 Capacidades .....	17
1.3 PROBLEMA U OPORTUNIDAD IDENTIFICADA .....	18
1.4 OBJETIVOS Y RESULTADOS ESPERADOS DEL PROYECTO .....	19
1.4.1 Objetivo General .....	19
1.4.2 Objetivos Específicos .....	19
1.4.3 Resultados Esperados .....	20
1.5 ALCANCE .....	21
1.6 RIESGOS POTENCIALES.....	22
<b>CAPÍTULO 2: MARCO TEÓRICO</b> .....	<b>23</b>
2.1 METODOLOGÍA DE INGENIERÍA DE NEGOCIOS .....	23
2.2 LÓGICAS DE NEGOCIO .....	26
2.2.1 Descubrimiento de Conocimiento en Base de Datos (KDD) .....	26
2.2.2 Clasificación Automática de Textos.....	31
<b>CAPÍTULO 3: PLANTEAMIENTO ESTRATÉGICO Y MODELO DE NEGOCIOS</b> .....	<b>38</b>
3.1 POSICIONAMIENTO ESTRATÉGICO .....	38
3.2 BALANCED SCORECARD .....	39

3.3 MODELO DE NEGOCIOS .....	40
3.3.1 Clientes .....	40
3.3.2 Propuesta de valor .....	40
3.3.3 Canales .....	41
3.3.4 Relación con los Clientes .....	42
3.3.5 Fuentes de Ingresos .....	42
3.3.6 Recursos Claves .....	42
3.3.7 Actividades Claves .....	43
3.3.8 Asociaciones Claves .....	43
3.3.9 Estructura de Costos .....	43
<b>CAPÍTULO 4: ANÁLISIS SITUACIÓN ACTUAL .....</b>	<b>44</b>
4.1 ARQUITECTURA DE PROCESOS .....	44
4.2 MODELAMIENTO DETALLADO DE PROCESOS .....	47
4.2.1 Modelamiento IDEF0 .....	47
4.2.2 Modelamiento BPMN .....	51
4.3 DIAGNÓSTICO DE LA SITUACIÓN ACTUAL .....	54
4.4 CUANTIFICACIÓN DEL PROBLEMA U OPORTUNIDAD .....	56
<b>CAPÍTULO 5: PROPUESTA DE DISEÑO DE PROCESOS .....</b>	<b>57</b>
5.1 DIRECCIONES DE CAMBIO Y ALCANCE .....	57
5.1.1 Estructura de la Empresa y Mercado .....	57
5.1.2 Asignación de Responsabilidades .....	57
5.1.3 Anticipación .....	57
5.1.4 Integración de Procesos Conexos .....	58
5.1.5 Mantenimiento Consolidado de Estados .....	59
5.1.6 Coordinación .....	59
5.1.7 Prácticas de Trabajo .....	60
5.1.8 Apoyo Computacional .....	60
5.2 DISEÑO DETALLADO DE PROCESOS TO BE .....	62

5.2.1 Administración Relación con los Sostenedores .....	62
5.2.2 Diseño en BPMN.....	70
5.3 DISEÑO DE LÓGICA DE NEGOCIOS .....	81
5.3.1 Sugerir Muestra a Fiscalizar en Programa Preventivo Uso de Recursos SEP .....	82
5.3.2 Asignación Óptima de Sostenedores y Escuelas .....	101
5.3.3 Decidir Acciones a las Recomendaciones de Gastos SEP .....	103
<b>CAPÍTULO 6: PROPUESTA DE APOYO TECNOLÓGICO .....</b>	<b>114</b>
6.1 ESPECIFICACIÓN DE REQUERIMIENTOS.....	115
6.1.1 Requerimientos Funcionales: Fichas de los Programas .....	115
6.1.2 Requerimientos Funcionales: Recomendaciones.....	117
6.1.3 Requerimientos No Funcionales.....	120
6.2 ARQUITECTURA TECNOLÓGICA .....	120
6.2.1 Arquitectura de Referencia.....	120
6.2.2 Framework y Patrones de Diseño .....	121
6.2.3 Diagrama de Despliegue y Componentes .....	122
6.3 DISEÑO DE LA APLICACIÓN .....	122
6.3.1 Casos de Uso: Fichas de los Programas.....	122
6.3.2 Casos de Uso: Recomendaciones .....	127
6.3.3 Diagramas de Secuencia: Fichas de los Programas .....	133
6.3.4 Diagramas de Secuencia: Recomendaciones .....	137
6.3.5 Diagramas de Clases: Ficha Programa .....	138
6.3.6 Diagramas de Clases: Recomendaciones.....	139
6.3.7 Prototipo Funcional Desarrollado: Ficha Programa .....	139
6.3.8 Prototipo Funcional Desarrollado: Recomendaciones .....	142
6.3.9 Datamart y Cubos OLAP .....	145
<b>CAPÍTULO 7: EVALUACIÓN DEL PROYECTO .....</b>	<b>149</b>
7.1 PLAN PILOTO .....	149
7.1.1 Definición del Plan Piloto.....	149

7.1.2 Resultados Obtenidos .....	149
7.2 DEFINICIÓN DE BENEFICIOS Y COSTOS.....	150
7.2.1 Beneficios .....	150
7.2.2 Costos.....	151
7.2.3 Inversiones.....	153
7.3 FLUJO DE CAJA .....	153
7.3.1 Depreciaciones .....	153
7.3.2 Tasa de Descuento Privada .....	154
7.4 ANÁLISIS DE SENSIBILIDAD .....	154
<b>CAPÍTULO 8: CONCLUSIONES .....</b>	<b>158</b>
8.1 SOBRE LA METODOLOGÍA DE INGENIERÍA DE NEGOCIOS.....	158
8.2 SOBRE EL CUMPLIMIENTO DE LOS OBJETIVOS DEL PROYECTO.....	158
8.3 SOBRE LA CONTINUIDAD DE LAS TÉCNICAS DATA MINING EN LA SUPEREDUC	160
8.4 SOBRE LAS TÉCNICAS UTILIZADAS PARA CONTRUIR EL CLASIFICADOR AUTOMATICO DE TEXTOS.....	161
8.5 SOBRE EL NUEVO SERVICIO DE RECOMENDACIONES A LAS RENDICIONES DE CUENTAS.....	162
8.6 TRABAJOS FUTUROS.....	163
<b>CAPÍTULO 9: BIBLIOGRAFÍA.....</b>	<b>164</b>
<b>CAPÍTULO 10: ANEXOS.....</b>	<b>168</b>
ANEXO N° 1: ESTRUCTURA ORGÁNICA Y DOTACIÓN DE PERSONAL .....	168
ANEXO N° 2: BALANCED SCORECARD DE LA SUPEREDUC .....	170
ANEXO N° 3: CONCENTRACIÓN DE VALORES EN VARIABLES .....	172
ANEXO N° 4: DETALLES ALGORITMO DE OPTIMIZACIÓN.....	186

## ÍNDICE DE TABLAS

TABLA N° 1: PRESUPUESTOS ANUALES DEL SECTOR EDUCACIONAL Y SUPEREDUC .....	14
TABLA N° 2: PRESUPUESTOS ANUALES EN TIC DEL SECTOR EDUCACIONAL Y SUPEREDUC .....	15
TABLA N° 3: VARIABLES Y VALORES .....	85
TABLA N° 4: RESULTADO TRATAMIENTO DATOS NULOS .....	85
TABLA N° 5: EXTRACTO DE MATRIZ DE CORRELACIÓN .....	87
TABLA N° 6: VARIABLES RESULTANTES DEL PROCESO DE LIMPIEZA.....	88
TABLA N° 7: VARIABLES A MODELOS ANALÍTICOS .....	89
TABLA N° 8: TABLA DE CENTROIDES DE CLÚSTERES .....	90
TABLA N° 9: GASTOS_RECHAZADOS VERSUS RETENCIONES.....	91
TABLA N° 10: GASTOS RECHAZADOS VERSUS MULTAS .....	91
TABLA N° 11: CARACTERIZACIÓN VERSUS PRIMER SET DE VARIABLES.....	92
TABLA N° 12: CARACTERIZACIÓN VERSUS SEGUNDO SET DE VARIABLES .....	92
TABLA N° 13: CARACTERIZACIÓN VERSUS TERCER SET DE VARIABLES .....	92
TABLA N° 14: MATRÍCULA POTENCIAL AFECTADA POR INCUMPLIMIENTOS .....	93
TABLA N° 15: EXTRACTO TABLA STOPWORD Y CORRECTOR ORTOGRÁFICO .....	106
TABLA N° 16: EXTRACTO DE TABLAS DE TÉRMINOS Y SUS VARIANTES .....	107
TABLA N° 17: EXTRACTO DE ARCHIVO DE SALIDA DEL CLASIFICADOR AUTOMÁTICO DE TEXTOS .....	112
TABLA N° 18: INCUMPLIMIENTOS SEP PROMEDIO ÚLTIMOS AÑOS .....	112
TABLA N° 19: RECOMENDACIONES CUENTA COMBUSTIBLE Y PEAJE RECURSOS AÑO 2016.....	113
TABLA N° 20: COMPARACIÓN RESULTADOS AÑOS PREVIOS VERSUS CLASIFICADOR .....	113
TABLA N° 21: PERFILES Y AUTORIZACIONES.....	117
TABLA N° 22: PROYECCIÓN DE BENEFICIOS DEL PROYECTO .....	150
TABLA N° 23: COSTO VISITA A TERRENO DE FISCALIZACIÓN.....	151
TABLA N° 24: ESTIMACIÓN DEL COSTO DE PERFILES SUPEREDUC .....	152
TABLA N° 25: DEPRECIACIONES DE LOS ACTIVOS DEL PROYECTO .....	153
TABLA N°26: FLUJO DE CAJA DEL PROYECTO .....	154
TABLA N° 27: FLUJO DE CAJA ESCENARIO SIN VISITAS DE FISCALIZACIÓN .....	155
TABLA N° 28: FLUJO DE CAJA ESCENARIO SIN ALGORITMO DE OPTIMIZACIÓN .....	156
TABLA N° 29: FLUJO DE CAJA PESIMISTA.....	157
TABLA N° 30: VALOR POLÍTICO.....	170
TABLA N° 31: VALOR BENEFICIARIO.....	170
TABLA N° 32: VALOR INTERNO .....	170
TABLA N° 33: VALOR FUTURO .....	171



## ÍNDICE DE ILUSTRACIONES

FIGURA N° 1: CAPACIDADES DE NEGOCIO DE LA SUPEREDUC .....	17
FIGURA N° 2: ONTOLOGÍA PARA EL DISEÑO DE NEGOCIO Y METODOLOGÍA DE INGENIERÍA DE NEGOCIO .....	23
FIGURA N° 3: ETAPAS DEL CRISP - DM.....	26
FIGURA N° 4: ANALÍTICA BASADA EN DATOS .....	27
FIGURA N° 5: ETAPAS DE LA CLASIFICACIÓN AUTOMÁTICA DE TEXTOS .....	32
FIGURA N° 6: POSICIONAMIENTO ESTRATÉGICO DE LA SUPEREDUC .....	38
FIGURA N° 7: CUADRO DE MANDO DE LA SUPEREDUC (SUPEREDUC, 2015) .....	39
FIGURA N° 8: ARQUITECTURA DE PROCESOS DE LA SUPEREDUC .....	44
FIGURA N° 9: FISCALIZACIÓN A SOSTENEDORES Y ESCUELAS .....	47
FIGURA N° 10: ADMINISTRACIÓN RELACIÓN CON LOS SOSTENEDORES Y ESCUELAS .....	49
FIGURA N° 11: ANÁLISIS DE SOSTENEDORES - ESCUELAS Y ACCIONES DE DIFUSIÓN NORMATIVA Y RECURSOS .....	50
FIGURA N° 12: GESTIÓN DE LAS FISCALIZACIONES.....	51
FIGURA N° 13: PREPARAR RANKING DE SOSTENEDORES Y/O ESCUELAS (AS IS) .....	52
FIGURA N° 14: OBTENER LA MUESTRA A FISCALIZAR (AS IS).....	53
FIGURA N° 15: ADMINISTRACIÓN RELACIÓN CON LOS SOSTENEDORES (TO BE) .....	62
FIGURA N° 16: ANÁLISIS DE SOSTENEDORES – ESCUELAS Y ACCIONES DIFUSIÓN NORMATIVA Y RECURSOS (TO BE) .....	63
FIGURA N° 17: ANALIZAR COMPORTAMIENTOS DE SOSTENEDORES Y ESCUELAS (TO BE) .....	64
FIGURA N° 18: DESARROLLAR MODELOS COMPORTAMIENTOS DE SOSTENEDORES Y ESCUELAS .....	66
FIGURA N° 19: DESARROLLAR MODELOS OBTENCIÓN MUESTRA Y RECOMENDACIONES PARA EL PROGRAMA.....	66
FIGURA N° 20: PLANIFICAR PROGRAMAS DE FISCALIZACIÓN (TO BE).....	68
FIGURA N° 21: PLANIFICAR PROGRAMAS DE FISCALIZACIÓN PREVENTIVOS .....	69
FIGURA N° 22: PREPARAR DATOS SOSTENEDORES Y ESCUELAS: FISCALIZACIÓN PREVENTIVA .....	71
FIGURA N° 23: PREPARAR DATOS SOSTENEDORES Y ESCUELAS: RECOMENDACIONES .....	72
FIGURA N° 24: EXPLORACIÓN DE DATOS .....	73
FIGURA N° 25: SELECCIÓN Y/O ADECUACIÓN DE VARIABLES A MODELOS .....	73
FIGURA N° 26: ELABORACIÓN DE MODELOS DE CLUSTERING SOSTENEDORES Y/O ESCUELAS .....	74
FIGURA N° 27: ELABORACIÓN DE MODELOS DE CLASIFICACIÓN SOSTENEDORES Y/O ESCUELAS.....	76
FIGURA N° 28: EVALUAR MODELOS COMPORTAMIENTOS DE SOSTENEDORES Y ESCUELAS .....	77
FIGURA N° 29: DESARROLLAR RECOMENDACIONES PARA EL PROGRAMA DE FISCALIZACIÓN .....	78

FIGURA N° 30: OBTENER MUESTRA A FISCALIZAR PARA EL PROGRAMA.....	79
FIGURA N° 31: PREPARAR RECOMENDACIONES AUTOMATIZADAS PARA EL PROGRAMA DE FISCALIZACIÓN.....	80
FIGURA N° 32: EXTRACTO DATAWAREHOUSE SUPEREDUC .....	84
FIGURA N° 33: CONCENTRACIÓN DE VALORES EN VARIABLE: CLASIFICACIÓN SEP.....	86
FIGURA N° 34: PROCESO DE IDENTIFICACIÓN DE VARIABLES CORRELACIONADAS.....	86
FIGURA N° 35: DETECCIÓN DE OUTLIERS MULTIVARIABLES.....	87
FIGURA N° 36: PROCESO DEL MODELO ANALÍTICO DE CLUSTERING .....	89
FIGURA N° 37: ÍNDICES DAVIES BOULDIN PARA DISTINTOS NÚMEROS DE CLÚSTERES.....	90
FIGURA N° 38: PROCESO DEL MODELO ANALÍTICO DE CLASIFICACIÓN .....	95
FIGURA N° 39: CARGA DE DATOS A MODELO DE CLASIFICACIÓN.....	95
FIGURA N° 40: FILTRAR SET DE DATOS POR CLÚSTER DE INTERÉS .....	96
FIGURA N° 41: PROCESO DE EQUILIBRADO DE LA VARIABLE OBJETIVO DEL MODELO.....	96
FIGURA N° 42: IMPORTANCIA DEL PREDICTOR .....	97
FIGURA N° 43: PROCESO DE PARTICIÓN DEL SET DE DATOS .....	97
FIGURA N° 44: MEJOR RESULTADO OBTENIDO CON ÁRBOLES CHAID .....	98
FIGURA N° 45: MEJOR RESULTADO OBTENIDO CON ÁRBOLES C5.0.....	98
FIGURA N° 46: RAMA ÁRBOL C5.0 SELECCIONADA.....	99
FIGURA N° 47: RAMA ÁRBOL C5.0, CONTINUACIÓN.....	100
FIGURA N° 48: RESULTADOS DEL ALGORITMO DE OPTIMIZACIÓN.....	103
FIGURA N° 49: ARCHIVO DE CARGA DE GASTOS Y SUS ATRIBUTOS.....	104
FIGURA N° 50: EXTRACTO DE DATOS .....	105
FIGURA N° 51: EXTRACTO RED SEMÁNTICA DE GRAFOS .....	107
FIGURA N° 52: RED SEMÁNTICA EN MOTOR DE BASE DE GRAFOS .....	108
FIGURA N° 53: ETAPAS DE LA CLASIFICACIÓN AUTOMÁTICA DE TEXTOS .....	108
FIGURA N° 54: EXTRACTO RED SEMÁNTICA USO DEBIDO, SEP Y CÓDIGO CUENTA = 410906.....	109
FIGURA N° 55: GRAFO DE RUTAS MÁS CORTAS PARA EL VECTOR DE TÉRMINOS .....	110
FIGURA N° 56: ARQUITECTURA DE REFERENCIA SUPEREDUC .....	121
FIGURA N° 57: DIAGRAMA DE DESPLIEGUE Y COMPONENTES.....	122
FIGURA N° 58: CASOS DE USO: MÓDULO FICHAS DE LOS PROGRAMAS FISCALIZACIÓN .....	126
FIGURA N° 59: CASOS DE USO: MÓDULO CLASIFICADOR DE TEXTOS - PREPROCESAMIENTO .....	131
FIGURA N° 60: CASOS DE USO: MÓDULO CLASIFICADOR DE TEXTOS – ENTRENAMIENTO .....	132
FIGURA N° 61: CASOS DE USO: MÓDULO CLASIFICADOR DE TEXTOS – CLASIFICADOR.....	132
FIGURA N° 62: DIAGRAMA SECUENCIA CU01: LOGIN.....	133
FIGURA N° 63: DIAGRAMA SECUENCIA CU02: CREAR/EDITAR FICHA PROGRAMA FISCALIZACIÓN ....	133

FIGURA N° 64: DIAGRAMA SECUENCIA CU03: ASOCIAR OBLIGACIONES NORMATIVAS .....	134
FIGURA N° 65: DIAGRAMA SECUENCIA CU04: REGISTRAR OBJETIVOS .....	134
FIGURA N° 66: DIAGRAMA SECUENCIA CU05: ADJUNTAR DOCUMENTOS.....	135
FIGURA N° 67: DIAGRAMA SECUENCIA CU06: CARGAR SUJETOS FISCALIZACIÓN.....	135
FIGURA N° 68: DIAGRAMA SECUENCIA CU07: INGRESAR CRITERIOS RIESGO .....	136
FIGURA N° 69: DIAGRAMA SECUENCIA CU08: CALCULAR RANKING .....	136
FIGURA N° 70: DIAGRAMA SECUENCIA CU19: CREAR NUEVO NODO TÉRMINO .....	137
FIGURA N° 71: DIAGRAMA SECUENCIA CU20: CREAR RELACIONES ENTRE TÉRMINOS .....	137
FIGURA N° 72: DIAGRAMA DE CLASES: FICHA PROGRAMAS .....	138
FIGURA N° 73: DIAGRAMA DE SECUENCIAS: RECOMENDACIONES .....	139
FIGURA N° 74: AUTENTICACIÓN.....	139
FIGURA N° 75: CREAR/EDITAR PROGRAMAS DE FISCALIZACIÓN: DATOS GENERALES .....	140
FIGURA N° 76: CREAR/EDITAR PROGRAMAS DE FISCALIZACIÓN: OBLIGACIONES NORMATIVAS .....	140
FIGURA N° 77: CREAR/EDITAR PROGRAMAS DE FISCALIZACIÓN: OBJETIVOS.....	140
FIGURA N° 78: CREAR/EDITAR PROGRAMAS DE FISCALIZACIÓN: CARGAR DOCUMENTOS.....	141
FIGURA N° 79: INGRESAR CRITERIOS DE RIESGO: CREAR CRITERIO.....	141
FIGURA N° 80: INGRESAR CRITERIOS DE RIESGO: CARGAR RESULTADOS POR CRITERIO .....	141
FIGURA N° 81: CALCULAR RANKING .....	141
FIGURA N° 82: ETL EXTRACCIÓN RENDICIÓN DE CUENTAS .....	142
FIGURA N° 83: CONTENIDO DEL ORIGEN DB UTILIZADO.....	142
FIGURA N° 84: ARCHIVO DE SALIDA PARA CLASIFICADOR AUTOMÁTICO DE TEXTOS .....	143
FIGURA N° 85: CARGAR ARCHIVO DE TEXTO A CLASIFICAR EN UNA CATEGORÍA.....	143
FIGURA N° 86: RESULTADOS DEL CLASIFICADOR TEXTOS PARA LA CATEGORÍA .....	143
FIGURA N° 87: MANTENEDOR DE NODOS DE LA RED SEMÁNTICA.....	144
FIGURA N° 88: MANTENEDOR DE RELACIONES DE LA RED SEMÁNTICA.....	144
FIGURA N° 89: PUBLICACIÓN DE RECOMENDACIONES: RENDICIÓN DE CUENTAS .....	144
FIGURA N° 90: CARGA DE RECOMENDACIONES EN SIFE .....	145
FIGURA N° 91: DATAMART Y CUBOS OLAP, PRIMERA PARTE .....	146
FIGURA N° 92: DATAMART Y CUBOS OLAP, SEGUNDA PARTE .....	146
FIGURA N° 93: EXTRACTO DE EXTRACCIÓN, TRANSFORMACIÓN Y CARGA DE DATOS.....	146
FIGURA N° 94: DETALLE DEL OPERADOR: CARGA OPERADOR HALLAZGO .....	147
FIGURA N° 95: EXTRACTO BASE DE CÁLCULO INDICADOR FECHA AGENDAMIENTO V/S EJECUCIÓN..	147
FIGURA N° 96: INDICADOR FECHA AGENDAMIENTO V/S EJECUCIÓN, ÚLTIMOS 45 DÍAS.....	148
FIGURA N° 97: SITIO DE GOBIERNO TRANSPARENTE DE LA SUPEREDUC .....	168
FIGURA N° 98: ORGANIGRAMA DE LA SUPEREDUC .....	169

FIGURA N° 99: DEPENDENCIA ADMINISTRATIVA ESCUELAS SEP .....	172
FIGURA N° 100: ZONA GEOGRÁFICA ESCUELAS SEP .....	172
FIGURA N° 101: CLASIFICACIÓN DE ESCUELAS SEP .....	173
FIGURA N° 102: EDUCACIÓN PARVULARIA EN ESCUELAS SEP .....	173
FIGURA N° 103: EDUCACIÓN BÁSICA EN ESCUELAS SEP .....	174
FIGURA N° 104: EDUCACIÓN MEDIA CIENTÍFICO HUMANISTA ESCUELAS SEP .....	174
FIGURA N° 105: EDUCACIÓN MEDIA TÉCNICO PROFESIONAL ESCUELAS SEP .....	175
FIGURA N° 106: MATRÍCULA AÑO 2016 ESCUELAS SEP .....	175
FIGURA N° 107: CANTIDAD DE CURSOS AÑO 2016 ESCUELAS SEP .....	176
FIGURA N° 108: MATRÍCULA PROMEDIO POR CURSO ESCUELAS SEP .....	176
FIGURA N° 109: GANANCIA DE MATRÍCULA ESCUELAS SEP, AÑO 2016 RESPECTO 2015.....	177
FIGURA N° 110: CANTIDAD DE DOCENTES EN ESCUELAS SEP .....	177
FIGURA N° 111: CANTIDAD DE ASISTENTES DE LA EDUCACIÓN EN ESCUELAS SEP .....	178
FIGURA N° 112: CANTIDAD DE DOCENTES Y ASISTENTES DE LA EDUCACIÓN EN ESCUELAS SEP ....	178
FIGURA N° 113: GANANCIA DE PLANTA ESCUELAS SEP, AÑO 2016 RESPECTO AÑO 2015 .....	179
FIGURA N° 114: TIPO DE APOYO EN GESTIÓN FINANCIERA ESCUELAS SEP .....	179
FIGURA N° 115: GASTO INCURRIDO EN PLANTA AÑO 2016 EN ESCUELAS SEP.....	180
FIGURA N° 116: INGRESOS SUBVENCIÓN NORMAL AÑO 2016 ESCUELAS SEP.....	180
FIGURA N° 117: RELACIÓN COSTO DE LA PLANTA VERSUS LOS INGRESOS EN ESCUELAS SEP .....	181
FIGURA N° 118: GASTOS RECHAZADOS AÑO 2014 EN ESCUELAS SEP .....	181
FIGURA N° 119: GASTOS RECHAZADOS AÑO 2015 EN ESCUELAS SEP .....	182
FIGURA N° 120: MULTAS AÑO 2014 ESCUELAS SEP .....	182
FIGURA N° 121: MULTAS AÑO 2015 ESCUELAS SEP .....	183
FIGURA N° 122: MULTAS AÑO 2016 ESCUELAS SEP .....	183
FIGURA N° 123: RETENCIONES AÑO 2014 EN ESCUELAS SEP .....	184
FIGURA N° 124: RETENCIONES AÑO 2015 EN ESCUELAS SEP .....	184
FIGURA N° 125: RETENCIONES AÑO 2016 EN ESCUELAS SEP .....	185

# **CAPÍTULO 1: INTRODUCCIÓN Y CONTEXTO**

## **1.1 ANTECEDENTES DE LA INDUSTRIA**

Para comprender el entorno en que opera una institución social, Bittar (2017) sugiere que es más apropiado utilizar el modelo PEST (Fahey & Narayanan, 1986), en vez del tradicional modelo de las cinco fuerzas de Porter.

### **1.1.1 Factor Político**

Chile, desde el retorno a la democracia en los años 90, cuenta con estabilidad política y un marco normativo en el sector educacional favorable para las instituciones estatales.

En el año 2006, producto del movimiento estudiantil conocido como la revolución pingüina, que reclama por las enormes brechas existentes en equidad y calidad en la educación, se presenta al Congreso Nacional el Proyecto de Ley General de Educación (LGE) y que, tras larga discusión, se aprueba (Ley General de Educación, N° 20.370, 2009). Esta Ley, deroga la LOCE (Ley Orgánica Constitucional de Enseñanza, N° 18.962, 1990) y se crea el Sistema de Aseguramiento de la Calidad de la Educación.

En el año 2011, producto de la Ley Sistema Nacional de Aseguramiento de la Calidad de la Educación Parvularia, Básica y Media y su Fiscalización, N° 20.529, 2011; se crea la Superintendencia de Educación, en adelante SUPEREDUC, y a la cual, en el año 2105 la Ley de Inclusión Escolar que Regula la Admisión de los y las estudiantes, Elimina el Financiamiento Compartido y Prohíbe el Lucro en Establecimientos Educativos que reciben Aportes del Estado, N° 20.845, 2015; le otorga mayores facultades y atribuciones.

### **1.1.2 Factor Económico**

La economía chilena tras su auge entre años 2010 - 2013, presenta una desaceleración producto al declive del precio del cobre y bajo consumo privado. A pesar de ello, el presupuesto anual del sector educacional crece año a año; lo anterior, se debe a la reforma educacional impulsada en el gobierno de Michelle Bachelet, y donde los recursos adicionales requeridos, provienen de la reforma tributaria implementada en el año 2015.

La siguiente tabla, extraída de Dirección de Presupuesto [DIPRES] (2016), resume el aumento presupuestario del sector educacional entre los años 2014 – 2017.

Tabla N° 1: Presupuestos anuales del sector educacional y SUPEREDUC

INSTITUCIONES	AÑO 2014 (M\$)	AÑO 2015 (M\$)	AÑO 2016 (M\$)	AÑO 2017 (M\$)
SUPERINTENDENCIA DE EDUCACIÓN	15.883.244	19.115.241	27.550.332	29.261.643
<b>TOTAL SECTOR EDUCACIÓN</b>	<b>7.189.064.757</b>	<b>8.082.014.548</b>	<b>9.379.446.986</b>	<b>10.325.235.930</b>

De la tabla anterior, se desprende el efecto de la reforma tributaria, y para el año 2017, el efecto de la estimación del 2,1% en el crecimiento económico, Banco Mundial (2016).

### **1.1.3 Factor Socio - Cultural**

En Chile la educación de calidad, gratuita y sin lucro es una preocupación social constante, al ser percibida como palanca de éxito para alcanzar estándares de vida superiores y al no confiar que las autoridades puedan resolver la problemática.

La población chilena, según Ministerio de Desarrollo Social [MDS] (2015) posee un nivel de escolaridad que alcanza los 10,8 años promedio y según el Instituto Nacional de Estadísticas [INE] (2012) la población no crece (envejece), es cada vez más urbana y el porcentaje de inmigrantes aumenta; además, según el Programa de Naciones Unidas para el Desarrollo [PNUD] (2012) en su informe sobre desarrollo humano en Chile, señala que la percepción de bienestar personal evoluciona positivamente desde la década de los 90, la educación aparece como palanca esencial para el éxito y la visión sobre lo público es cada vez más negativa; está última, según el Consejo para la Transparencia [CPT] (2016) en el año 2016 alcanzó niveles críticos, al señalar que 91% de los chilenos afirma que tiene poca o ninguna confianza que los políticos puedan solucionar los problemas del país, situación similar para parlamentarios, autoridades políticas y jueces.

### **1.1.4 Factor Tecnológico**

En Chile el gasto en TIC está muy por debajo de países desarrollados, a pesar de ello, el país lidera el aprovechamiento tecnológico a nivel regional.

Víctor Pérez (2011) señala, que contrariamente de lo que ocurre en países desarrollados, la actual institucionalidad que define las políticas de desarrollo científico, tecnológico y de

innovación en Chile, es frágil; lo cual, se condice con el gasto en TIC de Chile, que en año 2013 sólo alcanzó MM\$530.292, un 0,39% del PIB, y que según el Ministerio de Economía, Fomento y Turismo [MEFT] (2013), ubica a Chile en la última posición de un ranking compuesto por 29 países que encabeza Corea del Sur; a pesar de ello, Chile en el año 2014, es líder a nivel regional en aprovechamiento de las TIC, ocupando el puesto 35 a nivel global y donde Finlandia lidera el ranking (El Mostrador On Line [EMOL], 2014).

La siguiente tabla, extraída de Dirección de Presupuesto [DIPRES] (2016) resume el presupuesto anual en TIC del sector educacional entre los años 2014 – 2017

Tabla N° 2: Presupuestos anuales en TIC del sector educacional y SUPEREDUC

INSTITUCIONES	AÑO 2014 (M\$)	AÑO 2015 (M\$)	AÑO 2016 (M\$)	AÑO 2017 (M\$)
SUPERINTENDENCIA DE EDUCACIÓN	192.044	428.813	2.821.468	1.838.612
<b>TOTAL SECTOR EDUCACIÓN</b>	<b>7.207.997</b>	<b>8.551.487</b>	<b>14.476.471</b>	<b>11.033.521</b>

De la tabla anterior, se desprende el auge presupuestario en inversiones tecnológicas producto de la reforma tributaria y su caída en el año 2017, producto de las estimaciones pesimistas de crecimiento económico; independiente de esto, la SUPEREDUC presenta incrementos constantes en inversiones TIC, siendo una excepción respecto a todas las entidades fiscalizadoras del Estado (El Mercurio, 2017).

## 1.2 DESCRIPCIÓN GENERAL DE LA EMPRESA

La SUPEREDUC se crea en el año 2011 en base a la Ley Sistema Nacional de Aseguramiento de la Calidad de la Educación Parvularia, Básica y Media y su Fiscalización, N° 20.529, 2011; la cual, norma sus funciones y establece sus responsabilidades y obligaciones. La SUPEREDUC es un servicio público funcionalmente descentralizado, territorialmente desconcentrado de personalidad jurídica de patrimonio propio, se relaciona con el Presidente de la República por intermedio del Ministerio de Educación, en adelante MINEDUC, y su dotación alcanza los 750 funcionarios en el año 2017. Mayores detalles disponibles en el sitio Web [www.supereduc.cl](http://www.supereduc.cl) y en el Anexo N° 1: Estructura Orgánica y Dotación de Personal.

### **1.2.1 Misión**

Contribuir al aseguramiento de la calidad y el ejercicio del derecho a la educación parvularia y escolar, mediante la fiscalización del cumplimiento de la normativa, la rendición de cuentas, la gestión de denuncias y la provisión de información con un sentido de eficiencia, eficacia, transparencia y participación (SUPEREDUC, 2016).

### **1.2.2 Clientes**

Los clientes principales de la SUPEREDUC son los ciudadanos<sup>1</sup>, segmentados en: estudiantes, en los niveles de educación parvularia, básica y media; sus padres, apoderados o tutores; los docentes, asistentes de la educación, educadores y técnicos de párvulos; los directores y sostenedores de las escuelas; los ciudadanos comunes, que no pertenecen a los grupos anteriores; instituciones como: consejos escolares, centros de padres, instituciones del SAC<sup>2</sup>, agrupaciones (sostenedores, docentes, otras), Organismos del Estado, centros de investigación, fundaciones de educación, entre otras.

### **1.2.3 Servicios Principales**

Fiscalización a sostenedores y escuelas, dependiendo del hecho que las origine, se clasifican en planificadas y reactivas. Las fiscalizaciones reactivas provienen de las denuncias ciudadanas; mientras que las planificadas, responden al modelo de riesgos de la SUPEREDUC (actualmente en desarrollo). Las fiscalizaciones planificadas, a su vez, se clasifican en normativas y de uso de recursos; ambas, las normativas y de uso de recursos, se planifican anualmente, se ejecutan por medio de programas de fiscalización, se aplican a una muestra de sostenedores y escuelas, responden a una programación regional y a procedimientos coordinados de ejecución.

Atención ciudadana, recibe, investiga y resuelve denuncias, consultas y reclamos; y en determinados casos, se ofrece el servicio de mediación entre las partes en conflicto.

---

<sup>1</sup> Se refiere a un sentido amplio, al no estar restringida a la definición de ciudadano dada por la Constitución Política de la República de Chile en su artículo 13.

<sup>2</sup> Sistema Nacional de Aseguramiento de la Calidad (SAC) compuesto por: Ministerio de Educación, Agencia de la Calidad de la Educación, Consejo Nacional de Educación y la SUPEREDUC.



Promoción, formación y capacitación, instala conocimientos, capacidades, buenas prácticas y derechos educativos; todos, ad hoc a cada actor de la comunidad educativa.

Publicación de información, dispone información estadística y para la toma de decisiones de la ciudadanía; respecto de la ejecución de los distintos procesos de la SUPEREDUC.

### 1.2.4 Capacidades

El mapa de capacidades de negocio<sup>3</sup> ilustrado en la Figura N° 1, de elaboración propia y basado en los lineamientos de Ulrich & Rosen (2011), explicita como la SUPEREDUC brinda los servicios descritos y logra sus propósitos específicos y resultados.

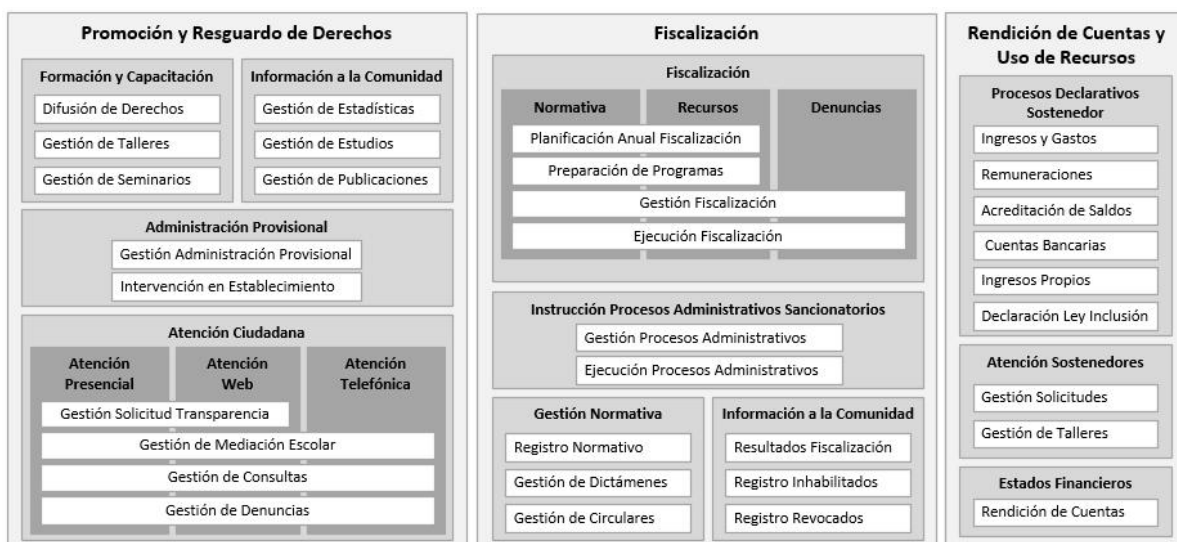


Figura N° 1: Capacidades de negocio de la SUPEREDUC

<sup>3</sup> Mapa de capacidades de alto nivel, elaborado en base a entrevistas con ejecutivos de la SUPEREDUC; no incluye capacidades de: dirección estratégica, planificación, gestión de compras, gestión de personas, gestión tecnológica.

### 1.3 PROBLEMA U OPORTUNIDAD IDENTIFICADA

La SUPEREDUC garantiza el cumplimiento de las condiciones necesarias para que exista calidad y equidad en la educación parvularia y escolar, previene incumplimientos normativos y vulneraciones de derechos educativos, y reintegra al sistema escolar aquellos recursos transferidos por el Estado que, tras ser fiscalizados, sean inexistentes, estén adulterados o no cuenten con la debida legalidad en su uso.

En la actualidad, de acuerdo al artículo 48 de la Ley N° 20.529, la SUPEREDUC debe fiscalizar el cumplimiento normativo educacional de más de 1.800 obligaciones y velar por la legalidad del uso de más de 6 billones de pesos anuales (3% presupuesto total del país) que reciben los sostenedores de establecimientos educacionales como aportes estatales. Para lo anterior, cuenta con una capacidad de fiscalización limitada, que promedia las 21.400 visitas anuales y que obliga a la Organización a priorizar las materias a incluir en cada programa, así como los sostenedores y/o escuelas a fiscalizar.

Los programas de uso de recursos fiscalizan la legalidad de los gastos y verifican la existencia de los bienes comprados y/o la prestación de los servicios contratados por los sostenedores. Esto se realiza en fecha posterior a la declaración anual<sup>4</sup> de gastos, por lo que durante el año no se toman medidas para evitar y/o mitigar riesgos derivados de usos indebidos de recursos. Con esta lógica operativa la SUPEREDUC anualmente detecta miles de millones de pesos en documentos riesgosos que debe fiscalizar, siendo en los fondos de la subvención escolar preferencial (SEP) -recursos que por Ley deben destinarse a apoyar a los estudiantes más vulnerables del país y a garantizar que los proyectos educativos se concreten- donde se concentra la mayor cantidad de incumplimientos.

En el año 2016, de la totalidad de fondos SEP transferidos por el Estado, la SUPEREDUC identificó la suma de MM\$ 76.000 en documentos riesgosos (representa un 9,5% del total de fondos SEP) que, tras ejecutar 2.100 visitas de fiscalización reintegraron sólo MM\$

---

<sup>4</sup> Los sostenedores de establecimientos educacionales subvencionados deben rendir anualmente cuenta del uso de todos sus recursos, respecto de la entidad sostenedora y de cada uno de sus establecimientos educacionales. Los recursos recibidos durante el año calendario anterior, se rendirán hasta el 31 de marzo del año siguiente (Art. 54 de la Ley N° 20.529).

22.250 (representa un 3% del total de fondos SEP) al sistema escolar. El rendimiento es menor si se compara la SUPEREDUC con otras instituciones que realizan labores similares, que detectan tasas de evasión<sup>5</sup> sobre 20% (sólo países europeos desarrollados declaran tener tasas menores a 10%).

En parte, el bajo rendimiento se explica en la información que respalda el uso de los gastos rendidos por los sostenedores al estar en formato semi - estructurado y donde, los actuales sistemas tienen dificultades para procesarla, integrarla y convertirla en información útil en los plazos<sup>6</sup> requeridos por los procesos de fiscalización; por consecuencia, esta información escasamente es utilizada en los actuales criterios de riesgo que detectan documentos riesgosos.

Por las razones expuestas, y sumado al principio de eficiencia del Estado, la SUPEREDUC debe aumentar la efectividad en las fiscalizaciones de uso de recursos y al mismo tiempo, ejecutar acciones preventivas y/o de mitigación para así, lograr disminuir año tras año el monto total en gastos no aceptados.

## **1.4 OBJETIVOS Y RESULTADOS ESPERADOS DEL PROYECTO**

### ***1.4.1 Objetivo General***

Preparar fiscalizaciones preventivas efectivas y focalizadas en aquellos sostenedores y escuelas que tienen mayor probabilidad de estar utilizando indebidamente los recursos SEP; de esta manera, optimizar los recursos de la SUPEREDUC al momento de preparar el programa de fiscalización y la planificación de las visitas a terreno.

### ***1.4.2 Objetivos Específicos***

- Segmentar y clasificar a sostenedores y escuelas en función de sus características y comportamientos relacionados con la gestión financiera, de modo que se tomen

---

<sup>5</sup> Fuentes: <http://www.uss.cl/economia-y-negocios/disminuyo-la-evasion-del-iva-chile-gracias-nuevo-modelo-fiscalizacion/> y [http://www.sii.cl/aprenda\\_sobre\\_impuestos/estudios/Antecede.htm](http://www.sii.cl/aprenda_sobre_impuestos/estudios/Antecede.htm)

<sup>6</sup> La Superintendencia no podrá aplicar ningún tipo de sanción luego de transcurridos seis meses desde la fecha en que hubiere terminado de cometerse el hecho (artículo 86 de la Ley N° 20.529).

acciones preventivas para disminuir usos indebidos de recursos SEP a partir del año 2018; utilizando para ello técnicas analíticas de data mining.

- Generar recomendaciones oportunas a los sostenedores que se encuentran realizando la rendición de cuentas SEP, de modo de evitar y/o mitigar usos indebidos de recursos y alertar errores de imputación de gastos a partir del proceso declarativo año 2018; utilizando para ello técnicas analíticas de data mining.
- Descubrir información en las descripciones de gastos de la rendición de cuentas SEP año 2017, para aumentar en al menos un 5% el monto en documentos riesgosos respecto del año anterior y fiscalizarlos a partir de octubre año 2018.
- Aumentar la actual capacidad de visitas de fiscalización para el programa uso de recursos SEP en un 2% respecto del año 2017. Esta capacidad podrá ser utilizada en el primer año para reintegrar recursos SEP al sistema escolar, en especial aquéllos provenientes de documentos riesgosos detectados en punto anterior.

### **1.4.3 Resultados Esperados**

Validar el rediseño del proceso de las fiscalizaciones para el programa uso de recursos SEP, sirviendo de guía para transitar desde una fiscalización reactiva a una cada vez más preventiva, que influya en el cambio de prácticas de los sostenedores al orientar y educar en el debido uso de los recursos SEP.

Obtener segmentos diferenciados de sostenedores y/o escuelas por riesgo de incumplimiento SEP, de tal forma que se puedan focalizar las acciones de fiscalización en aquéllos que presentan mayor riesgo de incumplimiento.

Diseñar, implementar y validar una solución tecnológica que procese la información semi-estructura contenida en las declaraciones de gastos SEP, que permita clasificar los gastos según uso (debido o indebido) e imputación contable (pertenece a la cuenta contable o no pertenece) para así, decidir si el documento riesgoso se incluye en la muestra a fiscalizar o se publica como recomendación en la plataforma Web.

Aumentar la capacidad de fiscalización bajo el supuesto de que existen tiempos ociosos que pueden ser utilizados al ser descubiertos por las actividades de seguimiento y control

de las fiscalizaciones al ser estas, apoyadas con herramientas del tipo *datamart*, cubos OLAP y *dashboards*.

Finalmente, que el proyecto sea un marco de trabajo de referencia para el diseño de procesos de fiscalizaciones preventivas en otros tipos de programas.

## 1.5 ALCANCE

- El proyecto se enmarca en las fiscalizaciones de uso de recursos, y de entre todos sus programas, se focaliza en el uso de recursos SEP; siendo su principal insumo, los gastos declarados y rendidos por los sostenedores adscritos al convenio SEP.
- De entre todos los procesos y/o actividades de la fiscalización de uso de recursos SEP, el proyecto se centra en el proceso de preparación de la muestra a fiscalizar para el programa. Este proceso es apoyado con analítica y con prácticas de trabajo para generar recomendaciones que eviten futuros gastos no aceptados al procesar la declaración de gastos mes a mes (no solo al final del año).
- Con respecto, a los procesos de gestión de las fiscalizaciones y de difusión de información, se deja propuesto el rediseño de procesos que establece las capacidades y configuraciones requeridas para operar.
- Es parte del proyecto diseñar y poblar una versión inicial de un *datawarehouse* (DW), con aquella información que permita lograr los objetivos del proyecto; siendo el diseño del DW escalable a otros programas de fiscalización.
- Es parte del proyecto diseñar e implementar una herramienta tecnológica que procese la información en formato semi-estructurado de las descripciones de gastos y prepare una base de datos ad hoc para aplicar analítica.
- Es parte del proyecto diseñar e implementar un algoritmo de programación matemática que apoye la decisión de asignación óptima de escuelas o sostenedores a incluir en la muestra de algún programa de fiscalización.
- Es parte del proyecto implementa un *datamart* e indicadores de seguimiento y control (cubos OLAP) utilizando la información transaccional del actual sistema TIC que soporta los procesos de gestión y ejecución de las fiscalizaciones.

- Queda fuera de alcance, los procesos y actividades relativos a la planificación regional, agendamientos y ejecución de las fiscalizaciones.
- Es parte del proyecto realizar las acciones de gestión del cambio que faciliten la adopción, materialización y/o dejar encaminadas las iniciativas propuestas.

## **1.6 RIESGOS POTENCIALES**

- De tipo organizacional, no lograr la efectividad esperada en los modelos analíticos y data mining, al no tener acceso a datos de fuentes externas (Impuestos Internos, MINEDUC, etc.); para mitigarlo, se tramitan de convenios de interoperabilidad.
- De tipo organizacional, escaso compromiso de los directivos y personal clave de la División de Fiscalización para implementar los trabajos planteados; es el principal riesgo del proyecto; para mitigarlo, se aplican técnicas de gestión del cambio (gestión de interesados, ofertas, negociación, etc.).
- De tipo cultural, resistencia al cambio de los fiscalizadores y coordinadores regionales al dar visibilidad del control del plan de ejecución de visitas de fiscalización; para mitigarlo, se debe trabajar en conjunto y hacerlos partícipes.

A los anteriores, y durante la etapa de implementación, se suman los siguientes:

- De tipo gestión y político, retrasos en la ejecución de las tareas planificadas debido a dependencia de personal clave y cargas de trabajos adicionales, producto de nuevas iniciativas no planificadas y priorizadas por la autoridad saliente debido al cambio de Gobierno (tiempos políticos); este riesgo se acepta, se replanifica el proyecto, y se informa a los interesados las nuevas fechas.
- De tipo político, cancelación o postergación del proyecto tras asumir las nuevas autoridades de Gobierno; este riesgo se previene aplicando técnicas de gestión del cambio y en particular, dando a conocer el proyecto con un discurso que resalte los intereses y preocupaciones de los nuevos interesados.

## CAPÍTULO 2: MARCO TEÓRICO

### 2.1 METODOLOGÍA DE INGENIERÍA DE NEGOCIOS

La ingeniería de negocios, según Barros (2015) no sólo comparte las ideas y principios de Service Science, Management, and Engineering (SSME), sino que también intenta cubrir un mayor dominio, incluyendo cualquier tipo de negocio; su énfasis es sobre cómo diseñar un negocio relacionando su estrategia, el modelo de negocio, capacidades implicadas, procesos y tecnologías de apoyo; en base a lo anterior, y experiencia en el diseño de muchas empresas diferentes, propone el modelo conceptual (ontología) para el diseño de un negocio (ver Figura N° 2), el cual da un marco de referencia al diseño.

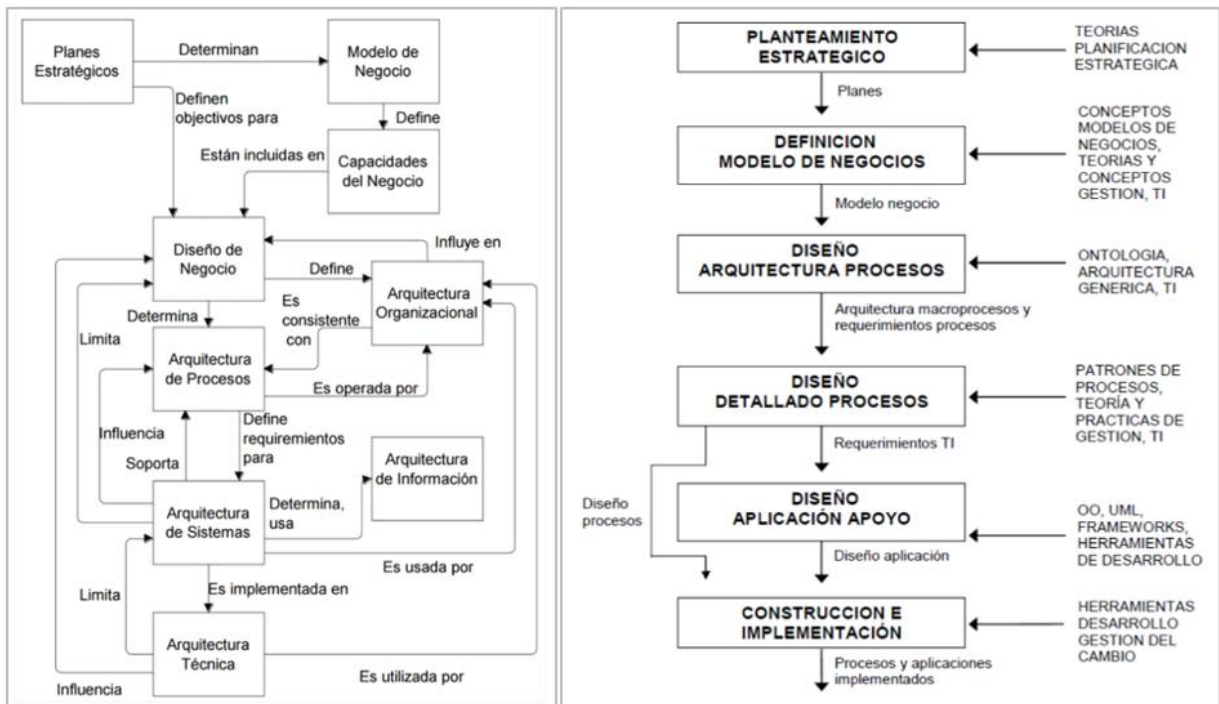


Figura N° 2: Ontología para el diseño de negocio y metodología de ingeniería de negocio

La metodología de ingeniería de negocios, sus etapas y orden secuencial para ser instanciada se ilustran en la Figura N° 2, y son descritas a continuación.

Planteamiento estratégico, según la Figura N° 2, los diseños se basan en los planes estratégicos y modelo de negocios que una organización quiere poner en marcha. El planteamiento estratégico, etapa inicial, asegura que el proyecto esté alineado con los objetivos estratégicos del negocio. El proyecto para definir el posicionamiento estratégico

utiliza el modelo delta de Arnoldo Hax (Hax, 2009) y para elaborar el cuadro de mando o *balanced scorecard* utiliza las directrices de Norton & Kaplan (2009).

Definición del modelo de negocios, los planes estratégicos determinan el modelo de negocio, el cual plantea como la organización crea valor a sus clientes y a otros *stakeholders*, recibiendo en retribución beneficios e ingresos. El proyecto para definir el modelo de negocios utiliza el modelo Canvas (Osterwalder & Pigneur, 2009).

Antes de continuar con las descripciones de la metodología, y fundamento para las siguientes etapas, es necesario conocer y precisar lo indicado por Barros (2015) “ninguna estrategia o modelo de negocio especifica cómo el posicionamiento y el valor serán realmente entregados en términos operacionales. Esto es lo que detallaremos en un diseño de negocio, a partir de capacidades de negocio necesarias según la estrategia y el modelo de negocio, lo cual debe ser complementado con el diseño de procesos, sistemas, organización y apoyo TI; que hacen las capacidades de negocio, plenamente operativas, dando lugar a las otras arquitecturas<sup>7</sup>” (p. 14).

Diseño de la arquitectura de procesos, establece los procesos necesarios para implementar las capacidades y diseño de negocio, las relaciones que coordinan los procesos, la lógica de negocio (algoritmos, heurísticas, reglas y, en general, procedimientos) que automatizan o guían dichos procesos y su conexión al soporte TI.

En la arquitectura de macroprocesos propuesta por Barros (2015), se definen cuatro agregaciones de procesos, que existen en cualquier organización, siendo estos:

- MACRO 1: cadena de valor, se define como un conjunto de procesos que ejecutan la producción de productos y servicios de una cierta línea de negocio de la organización. La cual, va desde la interacción con el cliente para generar requerimientos de tales productos y servicios hasta que son satisfactoriamente entregados.
- MACRO2: desarrollo de nuevas capacidades, se define como un conjunto de procesos que desarrollan las nuevas capacidades que la organización requiere

---

<sup>7</sup> Las “otras arquitecturas” se muestran en la Figura N° 2, las cuales conforman la arquitectura empresarial.



para ser competitiva, es donde se lleva a cabo la innovación requerida. La cual va desde: los nuevos productos y servicios, incluyendo los modelos de negocio, que la empresa requiere para mantenerse vigente en el mercado; la infraestructura necesaria para poder producir y operar los productos y servicios, incluyendo la infraestructura TI; y los nuevos procesos de negocios que aseguren la efectividad operacional y creación de valor para los clientes, estableciendo como consecuencia, los sistemas basados en tecnologías.

- MACRO 3: planificación del negocio, se define como un conjunto de procesos que establecen el curso futuro de la organización en la forma de estrategias, que se materializan en planes y programas.
- MACRO 4: procesos de apoyo (financieros, recursos humanos, materiales y otros), se define como un conjunto de procesos que manejan los recursos necesarios para que los macroprocesos anteriores, operen.

Los macroprocesos tienen relaciones genéricas entre ellos, y puede interpretarse como una arquitectura genérica que deben existir en cualquier organización y las relaciones que los ligan. Como tal, sirven como referencia al diseñar (instanciar) la arquitectura de procesos de un caso en particular.

Diseño detallado de procesos, en función de los patrones de procesos de negocios ad hoc para el caso en particular, se diseñan (instancian) los niveles de la arquitectura de procesos, hasta llegar a un detalle de bajo nivel (usando notación BPMN 2.0) en el que se identifican las actividades, prácticas de trabajo, lógicas y reglas de negocios que contribuyen a la implementación del modelo de negocio.

Diseño de la aplicación TI, a partir de los procesos definidos en la etapa anterior se diseñan las aplicaciones computacionales; utilizando para ello, notación UML (*Unified Modeling Lenguaje*), concepto de orientación a objetos, *frameworks*, otros.

Construcción e implementación, etapa final donde se escogen las tecnologías y se construyen (o adaptan) las aplicaciones (nuevas o existentes) diseñadas, las que están limitadas (o restringidas) por la arquitectura técnica de la organización.

## 2.2 LÓGICAS DE NEGOCIO

### 2.2.1 Descubrimiento de Conocimiento en Base de Datos (KDD)

El KDD según Fayyad, Piatetsky-Shapiro & Smyth (1996), permite por medio de un proceso no trivial identificar válidos, novedosos, potencialmente útiles y, comprensibles patrones ocultos y conexiones en entre los datos. En este contexto, según Hotho (2005) la expresión "potencialmente útiles" significa que los patrones encuentran una aplicación que genera un beneficio para el usuario; por lo tanto, la definición combina el descubrimiento del conocimiento con una aplicación específica.

El KDD es un proceso que se define en etapas, las que deben aplicarse a un conjunto de datos de interés para extraer patrones útiles; las etapas se realizan de manera iterativa, y varias de éstas generalmente requieren retroalimentación de un usuario. Según el modelo estándar definido por cross industry standard process for data mining (crisp - dm<sup>8</sup>) las principales etapas corresponden a: comprensión del problema o necesidad a resolver, comprensión de los datos, preparación de los datos, modelado, evaluación, e implementación; los dos primeros pasos se deben trabajar y comprender en conjunto, y uno de los pasos que más tiempo consume, es la preparación de datos.

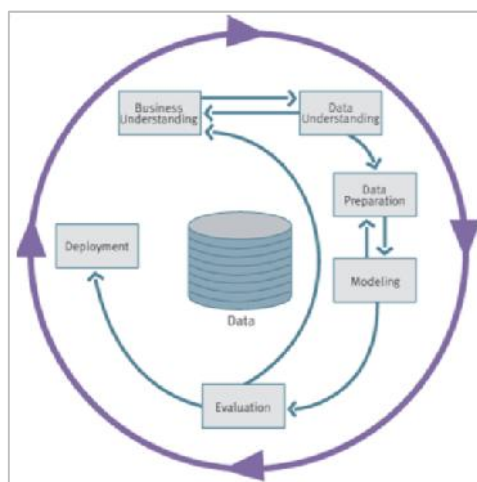


Figura N° 3: Etapas del crisp - dm

---

<sup>8</sup> <http://www.crisp-dm.org/>

Para el diseño de la solución tecnológica, escalable y reutilizable que provea información a los procesos de negocio respecto de: recomendaciones, automatizaciones o soporte; se utilizan las directrices dadas por Barros (2016) respecto a arquitecturas empresariales inteligentes, las cuales deben ser diseñadas a partir de la analítica o que la contengan.

En la Figura N° 4 se ilustra la analítica basada en datos según Davenport (2011), utilizada de referencia para el marco de trabajo del proyecto, y la cual se puede clasificar en:

- Analítica 1.0: base de datos analíticas, *datawarehouse* y visualización (*dashboards*).
- Analítica 2.0: *big data* interna, de la Web, redes sociales e Internet de las cosas (IoT), acompañada de analítica descriptiva.
- Analítica 3.0: todo lo anterior, más modelos predictivos (*mining* y *machine learning*) y prescriptivos (que establecen mejor curso de acción, incluidos los optimizantes).

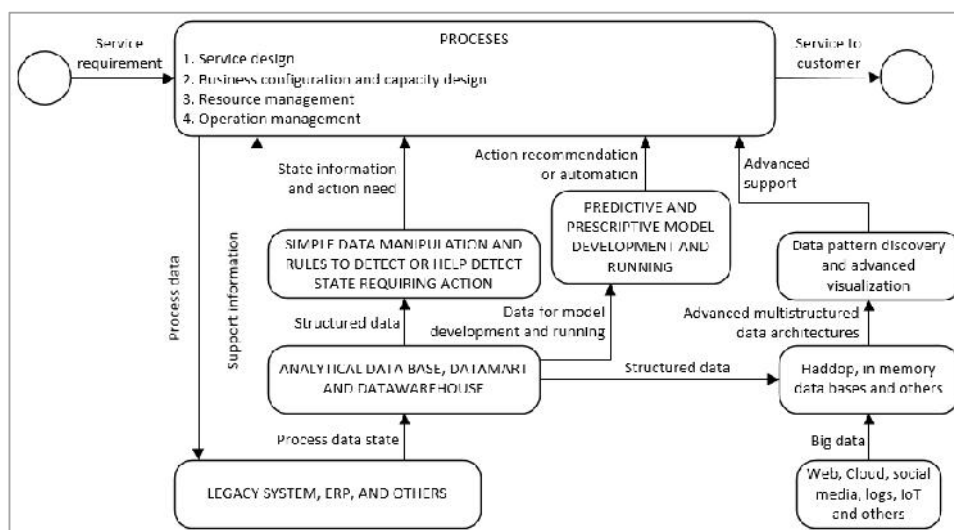


Figura N° 4: Analítica basada en datos

### 2.2.1.1 Clustering

*Clustering* según Han, Kamber & Pei (2012) es el proceso de particionar un conjunto de objetos de datos (observaciones) en subconjuntos, siendo cada subconjunto un clúster. Los objetos de un clúster son similares entre sí, pero diferentes a los objetos de otros clústeres. Diferentes métodos de agrupación pueden generar agrupaciones diferentes en el mismo conjunto de datos; siendo los principales métodos de agrupación:

particionamiento, jerárquicos, basados en densidad y basados en redes/grillas; de los métodos expuestos el proyecto utiliza el particionamiento, que es descrito a continuación.

#### **2.2.1.1.1 Métodos de Particionamiento**

Dado un conjunto de  $n$  objetos un método de partición construye  $k$  particiones de datos, donde cada partición representa un clúster o grupo ( $k \leq n$ ; cada grupo debe contener al menos un objeto). Los métodos de partición básicos adoptan típicamente la separación exclusiva de clústeres y están basados en distancia; siendo el criterio general de una buena partición, que los objetos de un grupo están cerca o relacionados entre sí, mientras que los objetos en otros clústeres estén muy separados o muy diferentes. De entre las distintas técnicas de partición disponibles, el proyecto utiliza *k-means*.

*K-means*: para un conjunto de datos  $D$  que contiene  $n$  objetos en un espacio euclidiano y tras indicar la cantidad  $k$  de particiones o clústeres, éstos son construidos  $C_1, \dots, C_k$  de forma tal que  $C_i \subset D$  y  $C_i \cap C_j = \emptyset$  para todo  $(1 \leq i, j \leq k)$ . El algoritmo inicia seleccionando aleatoriamente  $k$  objetos que serán los centroides  $c_i$  iniciales, cada objeto  $p$  restante se asigna a aquel clúster  $C_i$  más similar o que la distancia euclídea  $d(p, c_i)$  sea la menor (de no ser única, el objeto se asigna sólo a un clúster); ahora en cada iteración el algoritmo, recalcula un nuevo centroide  $c_i$  y reasigna cada objeto  $p$  a cada clúster  $C_i$ , las iteraciones continúan hasta que la asignación sea estable, es decir, los clústeres formados son los mismos que en la iteración anterior. La calidad del clúster  $C_i$  se mide por la variación intra clúster, que es la suma del error cuadrático  $E = \sum_{i=1}^k \sum_{p \in C_i} d(p, c_i)^2$  de entre todos los objetos de  $C_i$  y el centroide  $c_i$ . El método no garantiza converger al óptimo global y a menudo termina en un óptimo local. Para obtener buenos resultados en la práctica, es común ejecutar el algoritmo con diferentes centros de agrupamiento inicial.

Número óptimo de clústeres, a pesar de no existir un criterio objetivo o ampliamente validado, se utilizan con frecuencia estos tres métodos; el primer método elbow<sup>9</sup> (del

---

<sup>9</sup> EBK-Means: A Clustering Technique based on Elbow Method and K-Means in WSN, disponible en <http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.735.7337&rep=rep1&type=pdf>

codigo), utiliza los valores de la inercia obtenida tras aplicar el *k-means* a diferente número de clústeres, siendo la  $I_1 = \sum_{i=0}^n ||x_i - \mu||^2$  la suma de las distancias al cuadrado de cada objeto del clúster a su centroide; los valores obtenidos se representan en una gráfica lineal (inercia v/s número de clústeres) el punto en el que se observa el cambio brusco en la inercia, será el número óptimo de clústeres. El segundo método índice de Davies Bouldin, se calcula para cada clúster en función de la distancia intra-clúster e extra-clústeres, y se consideran el mejor número de clústeres aquel que promedia el menor índice. El tercer método, utiliza el promedio de la distancia intra - clúster respecto a su centroide (similar al anterior).

### **2.2.1.2 Clasificación**

Clasificación según Han, Kamber & Pei (2012) es un proceso que consta de una etapa de aprendizaje, donde se construye el modelo con una parte de los datos, y una etapa de predicción, donde el modelo predice para el otro conjunto de datos las etiquetas de clases; en la primera etapa, el algoritmo construye el clasificador, analizando o aprendiendo de un conjunto aleatorio de tuplas de datos donde cada tupla  $X$  representa un vector de atributos  $n$ -dimensional  $X = \{x_1, \dots, x_n\}$  y cada atributo es una característica de  $X$  y cada tupla pertenece a una clase predefinida, cuando la precisión del clasificador se considera aceptable se utiliza para clasificar tuplas de datos para las cuales las etiquetas de clases se desconocen. Los principales métodos de clasificación son: árboles de decisión, redes neuronales artificiales, support vector machines. De entre las distintas técnicas de clasificación disponibles, el proyecto utiliza árboles de decisión.

#### **2.2.1.2.1 Árboles de Decisión**

Es un diagrama de flujo con estructura de árbol; donde, cada nodo interno representa un test de un atributo y las ramas representan los resultados del test, y cada nodo hoja (o nodo terminal) tiene una etiqueta de clase, siendo el nodo superior en un árbol, la raíz. Algunos algoritmos de árboles de decisión son: ID3, C4.5 y CART; este último, produce solo árboles binarios, donde cada nodo interno se ramifica solo en otros dos nodos. A modo general, el algoritmo básico de árboles de decisión es:

- Se crea un árbol inicial con un único nodo  $n$ , sin etiquetar y se asigna como conjunto de ejemplos todo el conjunto de entrenamiento.

- Se iterar como sigue:
  - Primer nodo  $n$  sin etiquetar, si los ejemplos asignados tienen todos la misma clase, se etiqueta  $n$  con esa clase; en caso contrario se aplican criterios de separación, que se describen en lo sucesivo.
  - Etiquetar  $n$  con el mejor atributo  $a$ , según el conjunto de ejemplos asignados; para cada valor de  $a$  se crea una nueva arista descendente en el nodo  $n$  y al final de cada una de esas nuevas aristas se crea un nuevo nodo sin etiquetar,  $n_1, \dots, n_k$ ; luego, se separan los ejemplos asignados al nodo  $n$  según el valor que tomen para el atributo  $a$  y se crean nuevos conjuntos de ejemplos para  $n_1, \dots, n_k$ .
  - Iterar hasta que todos los nodos estén etiquetados.

A continuación, se describen tres criterios de separación o medidas de selección de atributos: *information gain*, *gain ratio*, y *gini index*; utilizando para ello, la siguiente notación: conjunto  $D$ , datos de la partición del conjunto tuplas de clases-etiquetadas de entrenamiento. Cada atributo de la clase-etiquetada tiene  $m$  distintos valores, definiendo  $m$  distintas clases  $C_i$  (para cada  $i = 1, \dots, m$ ). El conjunto  $C_{i,D}$  son las tuplas de las clases  $C_i$  en  $D$ .  $|D|$  y  $|C_{i,D}|$  que denota el número de tuplas en  $D$  y  $C_{i,D}$  respectivamente.

*Information gain*, el algoritmo ID3 usa esta medida de selección de atributos, el nodo  $n$  contiene las tuplas de la partición  $D$ . El atributo con la mayor ganancia de información se elige como el atributo de división para el nodo  $n$ . Este atributo minimiza la información necesaria para clasificar las tuplas en particiones resultantes y refleja la menor aleatoriedad o impureza en estas particiones. Tal enfoque, minimiza el número esperado de pruebas necesarias para clasificar una tupla dada y garantiza un árbol simple. La información esperada necesaria para clasificar una tupla en  $D$  viene dada por  $I_1(D) = -\sum_{i=1}^m p_i \log_2(p_i)$ ; donde  $p_i$  es la probabilidad distinta de cero que una tupla arbitraria en  $D$  pertenezca a la clase  $C_i$  y es estimada por  $|C_{i,D}| / |D|$ . Ahora, se debe dividir las tuplas en  $D$  en algún atributo  $a$  que tenga valores distintos  $\{a_1, \dots, a_v\}$  observados del conjunto de entrenamiento, si  $a$  tiene un valor discreto estos valores corresponden directamente a los resultados  $v$  de una prueba en  $a$ ; el atributo  $a$  puede ser utilizado para dividir  $D$  en  $v$

particiones o subconjuntos  $\{D_1, \dots, D_v\}$ , idealmente esta partición debería producir una clasificación exacta de las tuplas (cada partición sea pura, bastante improbable); siendo  $I_{i, a}(D) = \sum_{j=1}^v \frac{|D_j|}{|D|} \times i_{i, a}(D_j)$  la cantidad de información que se necesita, después de la partición, para una clasificación exacta. La *information gain* se define como  $G(a) = I_{i, a}(D) - I_{i, a}(D)$ , el atributo  $a$  con la mayor *information gain* se elige como el divisor en el nodo  $n$ .

*Gain ratio*, el algoritmo C4.5 usa esta medida de selección de atributos, e intenta superar el sesgo de la medida anterior cuando se enfrenta a pruebas con muchos resultados. Esta medida se aplica a un tipo de información normalizada definida por  $S_{i, a}(D) = -\sum_{j=1}^v \frac{|D_j|}{|D|} \times \log_2\left(\frac{|D_j|}{|D|}\right)$ , este valor representa la información potencial generada al dividir el conjunto de entrenamiento de datos  $D$  en particiones  $v$  correspondientes a los resultados  $v$  de una prueba en el atributo  $a$ ; para cada resultado, considera el número de tuplas que tienen ese resultado con respecto al número total de tuplas en  $D$ . La relación de ganancia se define como  $G(a) = \frac{G(a)}{S_{i, a}(D)}$ , el atributo  $a$  con la relación de ganancia máxima se elige como el divisor en el nodo  $n$ .

*Gini index*, el algoritmo CART usa esta medida de selección de atributos, y mide la impureza de  $D$ , una partición de datos o un conjunto de tuplas de entrenamiento, como  $G(D) = 1 - \sum_{i=1}^m p_i^2$ ; donde  $p_i$  es la probabilidad de que una tupla en  $D$  pertenezca a la clase  $C_i$  y es estimada por  $|C_{i,D}| / |D|$ . El *gini index* considera una división binaria para cada atributo, siendo el atributo que maximiza la reducción en la impureza (o menor índice) se elige como el divisor en el nodo  $n$ .

## 2.2.2 Clasificación Automática de Textos

Según Baeza-Yates & Ribeiro-Neto (1996) la clasificación automática de textos es un proceso supervisado que requiere de un conjunto de documentos previamente clasificados por expertos humanos que funcionan como entrenamiento para el sistema y donde, según Mitchell (1997) tiene por objeto asignar clases predefinidas a documentos de texto (ver Figura N° 5). La tarea de clasificación comienza con un conjunto de documentos de entrenamiento  $D = (d_1, \dots, d_m)$  etiquetados con una (o más) clase(s)  $C_i$ , con ello se determina el método de clasificación capaz de asignar la clase correcta a un

nuevo documento  $d$  del dominio; para determinar el rendimiento del método se utilizan las métricas de  $p$  (e. ) =  $\frac{v}{v \epsilon} \frac{p}{p} \frac{it}{+f p}$  y  $r$  (m ) =  $\frac{v}{v} \frac{p}{p} \frac{+f}{+f} \frac{n}{n}$ , donde la “alta precisión” significa que el método arroja resultados sustancialmente más relevantes que irrelevantes, mientras que un alto recall significa que el método devolvió la mayoría de los resultados relevantes; obviamente, hay una compensación entre la precisión y el recall, siendo la métrica del desempeño general del método  $F = \frac{2}{((1/r) + (1/p))}$ .

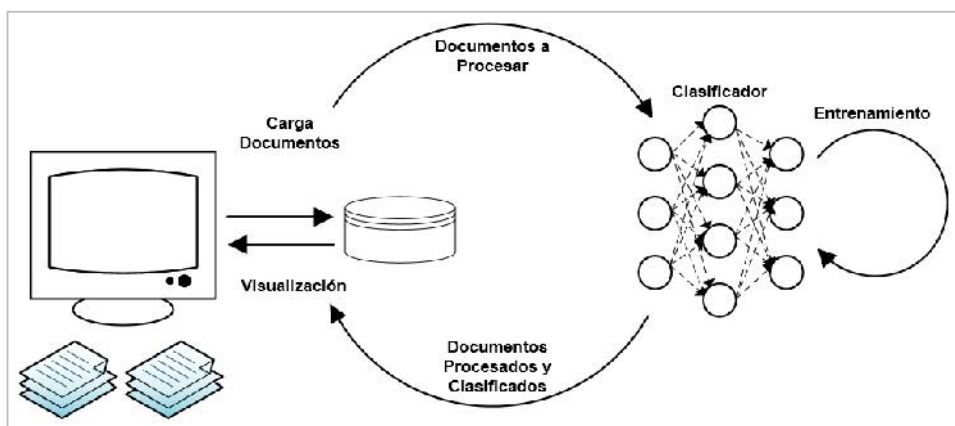


Figura N° 5: Etapas de la clasificación automática de textos

En base al trabajo de Cárdenas, Olivares & Alfaro (2014) que compara distintos clasificadores para textos cortos (hasta 500 palabras), selecciona por desempeño general un *clasificador basado en redes de palabras*, que es descrito en lo sucesivo. Siendo este trabajo símil a la problemática que el proyecto va a resolver. Para entender este clasificador, es requisito conocer las técnicas de *text mining* y teoría de grafos.

### 2.2.2.1 Text Mining

Según Hotho (2005) *text mining*, necesita métodos especiales de pre procesamiento para convertir datos textuales en un formato adecuado para aplicar los algoritmos, los que frecuentemente requieren del desarrollo de nuevos o la adaptación de algoritmos. *Text mining* aborda los problemas de representación de textos, clasificación, agrupamiento, extracción de información o búsqueda y modelado de patrones ocultos.



Para extraer grandes colecciones de documentos, es necesario pre procesarlos y almacenarlos en una estructura de datos apropiada (vector de palabras) para un procesamiento posterior, más que sólo un archivo de texto sin formato; la representación vectorial permite almacenar el valor numérico de importancia de cada palabra. Los enfoques actualmente predominantes basados en esta idea son el modelo de espacio vectorial (Salton, Wong & Yang, 1975), el modelo probabilístico (Robertson, 1977) y el modelo lógico (van Rijsbergen, 1986).

Preprocesamiento de texto, para obtener todas las palabras que se usan de un texto dado se requiere un proceso de tokenización, donde un texto se divide en una secuencia de palabras eliminando toda la puntuación, marcas y otros caracteres que no sean de utilidad; siendo el diccionario, el conjunto de palabras diferentes obtenida al fusionar todos los documentos de texto de una colección.

Filtrado, lematización y *stemming*, son utilizadas para reducir el tamaño del diccionario y, por lo tanto, la dimensionalidad de la descripción de documentos dentro de la colección. El filtrado consiste en eliminar palabras que contienen poca o ninguna información de contenido, como artículos, conjunciones, preposiciones, etc.; además, según Frakes & Baeza-Yates (1992) las palabras que ocurren extremadamente a menudo poseen poco contenido de información para distinguir entre documentos, y también palabras que ocurren muy pocas veces es probable que no tengan relevancia estadística, por ello, pueden eliminarse del diccionario. La lematización consiste en hallar el lema o forma que por convenio se acepta como representante de todas las formas flexionadas de una misma palabra (puede ser morfológica o sintáctica). El *stemming* construye las formas básicas de las palabras, es decir, eliminar el plural de sustantivos u otras afinidades.

Selección términos índices, para disminuir aún más el número de palabras que se deben usar, es posible indexar o seleccionar palabras claves que describen los documentos; un método simple para seleccionar palabras claves se basan en su entropía (Lochbaum & Streeter, 1989), donde para cada palabra  $t$ :

$$W(t) = 1 + \frac{1}{\log_2 |D|} \sum_{d \in D} P(d, t) \log_2 P(d, t) \text{ con } P(d, t) = \frac{t(d, t)}{\sum_{t=1}^n t(d, t)}$$

Aquí la entropía da una medida de qué tan bien una palabra es adecuada para separar documentos por palabra clave de búsqueda; por ejemplo, las palabras que aparecen en muchos documentos tendrán poca entropía.

Modelo de espacio vectorial, representa los documentos como vectores en el espacio m-dimensional, es decir, cada documento  $d$  se describe mediante un vector de características numéricas  $w(d) = (x(d, t_1), \dots, x(d, t_m))$ ; por lo tanto, los documentos se pueden comparar mediante el uso de operaciones vectoriales simples, cada elemento del vector representa una palabra (o grupo de palabras). La codificación consiste en utilizar vectores de términos binarios; es decir, un elemento vectorial se establece en 1, si la palabra se usa en el documento o 0 si no; esta codificación asigna similar importancia a todos los términos para una consulta o comparación específica, para mejorar los rendimientos se ponderan los términos, los pesos reflejan la importancia de una palabra en un documento específico, pesos grandes se asignan a términos que se usan con frecuencia en documentos relevantes, pero rara vez en toda la colección considerada. Por lo tanto, peso  $w(d, t)$  para un término  $t$  en el documento  $d$  se calcula por frecuencia de término  $t$ ,  $f(d, t)$  y la frecuencia inversa del documento  $if(t)$  (donde  $if(t) = \log\left(\frac{N}{n_t}\right)$ ).

$$w(d, t) = \frac{f(d, t) \cdot if(t)}{\sqrt{\sum_{j=1}^m f(d, t_j)^2 \cdot (if(t_j))^2}}$$

Donde  $N$  es el tamaño de documentos de la colección  $D$  y  $n_t$  es el número de documentos en  $D$  que contiene el término  $t$ .

Sobre un esquema de ponderadores, un documento  $d$  se define por el vector de ponderaciones  $w(d) = (x(d, t_1), \dots, x(d, t_m))$  y la similitud  $S$  de dos documentos  $d_1$  y  $d_2$  se puede calcular en función del producto de los vectores (si se asume vectores normalizados es posible calcular el coseno entre los dos vectores).

$$S(d_1, d_2) = \sum_{k=1}^m w(d_1, t_k) \cdot w(d_2, t_k)$$

Una medida de distancia frecuente es la distancia euclidiana, la que solo debe usarse para vectores normalizados; de lo contrario, se debe utilizar la similitud coseno.

Similitud coseno, un documento  $d$  se puede representar por miles de términos  $t$ , cada uno grabando la frecuencia de ocurrencia o si codificación binaria (1 si el término, palabra o frase se usa en el documento o 0 si no), los vectores de frecuencia de término suelen ser muy largos y escasos (es decir, tienen muchos valores 0); según Han, Kamber & Pei (2012) las medidas de distancias tradicionales no funcionan bien para un número tan escaso de datos, en otras palabras, se necesita una medida para los datos numéricos que ignore las coincidencias de ceros; por ello, se utiliza la similitud coseno  $si(x, y) = \frac{x \cdot y}{||x|| ||y||}$ ;

donde  $||x||$  corresponde a  $\sqrt{x_1^2 + \dots + x_p^2}$  la longitud del vector (de igual forma para  $y$ ). La medida calcula el coseno del ángulo entre vectores  $x$  e  $y$ . Un valor de coseno de 0 significa que los dos vectores están a 90 grados uno del otro (ortogonales) y no tienen coincidencias; cuanto más cerca esté el coseno de 1, menor ángulo y mayor es la coincidencia entre vectores.

Preprocesamiento lingüístico, a menudo, los métodos de minería de textos pueden aplicarse sin más preprocesamiento. A veces, sin embargo, se puede usar preprocesamiento lingüístico adicional para mejorar la información disponible sobre los términos; como, por ejemplo, marcas para verbos, adjetivos, sustantivos, otro; agrupar palabras adyacentes en una oración; resolver la ambigüedad en el significado de palabras sueltas o frases; análisis completo de una oración, para encontrar la relación de cada palabra en la oración con todas las demás, y típicamente también su función en la oración (sujeto, objeto, otra) u otras reglas manuales.

### **2.2.2.2 Clasificador Basado en Redes de Palabras**

Previo a describir este clasificador es necesario describir sucintamente la teoría de grafos. El siguiente extracto se obtiene del trabajo de Graham, Lenstra & Tarjan (1999).

Extracto teoría de grafos, un grafo  $G$  es un par ordenado  $G = \{V, E\}$ ; donde  $V$  es un conjunto de vértices o nodos y  $E$  un conjunto de aristas o arcos que relacionan los nodos. Se llama orden del grafo  $G$  a su número de vértices  $|V|$  y los grafos se clasifican en no dirigidos, dirigidos, mixtos, valorados, entre otros.

Un grafo dirigido es un grafo  $G = \{V, E\}$  donde:  $V$  ;  $E = \{(v_i, v_j) \mid v_i, v_j \in V\}$  es un conjunto de pares ordenados de elementos de  $V$ ; dada una arista  $(v_i, v_j)$ ,  $v_i$  es su nodo inicial y  $v_j$  su nodo final; por definición los grafos dirigidos no tienen bucles (una arista donde el nodo inicial y el nodo final coinciden); y se denomina valorado, cuando las aristas tienen un valor numérico.

Las propiedades de los grafos son: adyacencia, dos aristas son adyacentes si tienen un vértice en común, y dos vértices son adyacentes si una arista los une; incidencia, una arista es incidente a un vértice si ésta lo une a otro; ponderación, corresponde a una función que a cada arista le asocia un valor (costo, peso, longitud, etc.), para aumentar la expresividad del modelo. Esto se usa mucho para problemas de optimización, como el del vendedor viajero o del camino más corto; y etiquetado, distinción a los vértices y/o aristas mediante una marca que los hace unívocamente distinguibles del resto.

Para representar los grafos se utiliza la matriz de adyacencia, donde las filas y columnas son los vértices, y en cada  $(v_i, v_j)$  pone un 0 para indicar adyacencia o un 1 si la hay.

Un recorrido en un grafo  $G = \{V, E\}$ , es una sucesión de vértices  $v_0, v_1, \dots, v_k$  tal que,  $\{v_i, v_{i+1}\} \in E$  para todo  $0 \leq i < k$ ; la longitud de un recorrido  $v_0, v_1, \dots, v_k$  es  $k$ . Un recorrido puede repetir vértices, y puede comenzar y acabar en vértices diferentes.

Un camino es un recorrido  $v_0, v_1, \dots, v_k$  en el que  $v_i \neq v_j$  para todo  $0 \leq i, j < k$  con  $i \neq j$ ; es decir, en un camino todos los vértices son distintos entre sí, excepto quizás el primero y el último.

Existen distintos tipos de algoritmos de búsquedas, como son: búsqueda en anchura (BFS), búsqueda en profundidad (DFS), del vecino más cercano, Dijkstra, Bellman-Ford, Prim, Ford-Fulkerson, Kruskal y Floyd-Warshall; este último, será utilizado por el proyecto para las redes de palabras y encuentra el camino mínimo en grafos dirigidos ponderados, en una única ejecución.

Redes de palabras, se define un grafo dirigido  $G_i$  como aquella red de palabras construida a partir de un conjunto de textos clasificados en la categoría  $i$  por un humano; notar que habrá tantos grafos como categorías tenga el problema a resolver). El texto a ser

clasificado se define en otro grafo dirigido  $g$  (el texto asociado a  $g$  y aquellos que formaron  $G_i$  fueron previamente pre procesado, tal como fue descrito en sección anterior).

El proceso de clasificación sigue el procedimiento: si el grafo  $g$  es un subgrafo completo o parcial respecto de la cantidad de vértices de  $g$  en  $G_i$  (ídem procedimiento en cada categoría  $i$ ); en caso de que  $g$  sea subgrafo en más de una categoría, se define una función que sirva de medida de decisión respecto de la categoría a ser asignada a  $g$ .

Cárdenas, Olivares & Alfaro (2014) definen un costo de construcción  $C$  para un texto  $h$  en una determinada categoría como,  $C = \left(\frac{l^\alpha}{P^\beta}\right) + v$ ; donde,  $l$  es el promedio de las distancias más cortas entre todos los pares de palabras co-ocurrentes de  $h$  presentes en  $G$  (número de nodos mínimo que debe recorrerse en  $G$  para unir cada par),  $P$  la cantidad de palabras de  $h$  presentes en  $G$ ,  $\alpha$  y  $\beta$  las ponderaciones de ambas medidas y  $v$  un valor (negativo) obtenido a partir de una valoración de las palabras y símbolos presentes en  $h$ ; de esta forma, mientras más palabras de  $h$  estén en  $G$ , y menor sea la distancia entre éstas en  $G$ , menor será el costo de construcción de  $h$  en  $G$ .

## CAPÍTULO 3: PLANTEAMIENTO ESTRATÉGICO Y MODELO DE NEGOCIOS

### 3.1 POSICIONAMIENTO ESTRATÉGICO

Se utiliza el Modelo Delta de Hax (Hax, 2009) para determinar el posicionamiento estratégico de la SUPEREDUC. Como ilustra la Figura N° 6, el posicionamiento estratégico se encuentra ubicado en la arista soluciones al cliente y consolidación del sistema, específicamente en la posición de canal exclusivo; siendo las razones:

- Ley Sistema Nacional de Aseguramiento de la Calidad de la Educación Parvularia, Básica y Media y su Fiscalización, N° 20.529, 2011; la crea como ente fiscalizador y cuenta con una casi completa consolidación del sistema. Sólo la Contraloría General de la República puede auditar el uso de recursos a sostenedores municipales, que en año 2017 corresponden a 354 (en total son más de 5.600).
- La Ley de Inclusión Escolar que Regula la Admisión de los y las Estudiantes, Elimina el Financiamiento Compartido y Prohíbe el Lucro en Establecimientos Educativos que reciben Aportes del Estado, N° 20.845, 2015; le otorga obligaciones adicionales, mayores facultades y atribuciones.

El desafío de la SUPEREDUC, a la inversa de las organizaciones con fines de lucro, es bajar por la arista consolidación del sistema y mejor producto, creando soluciones diferenciadoras; luego, moverse hacia solución integral al cliente, redefiniendo la experiencia con el cliente; para luego, lograr de manera justa, reconocida y valorada la consolidación del sistema.

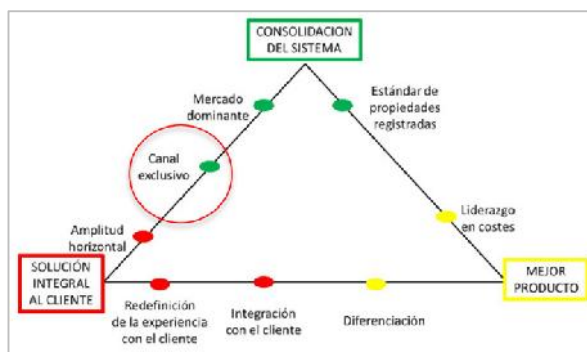


Figura N° 6: Posicionamiento estratégico de la SUPEREDUC

### 3.2 BALANCED SCORECARD

Se utilizan las directrices de Norton & Kaplan (2009) para elaborar el cuadro de mando de la SUPEREDUC y sus recomendaciones para lograr el éxito en las organizaciones sin ánimo de lucro o gubernamentales; al señalar que estas, “deben medirse por cuán eficiente y eficazmente satisfacen las necesidades de los usuarios, y los objetivos tangibles deben ser definidos por los clientes y usuarios”. La Figura N° 7 ilustra el cuadro de mando de la SUPEREDUC, sus detalles se presentan en el Anexo N° 2.

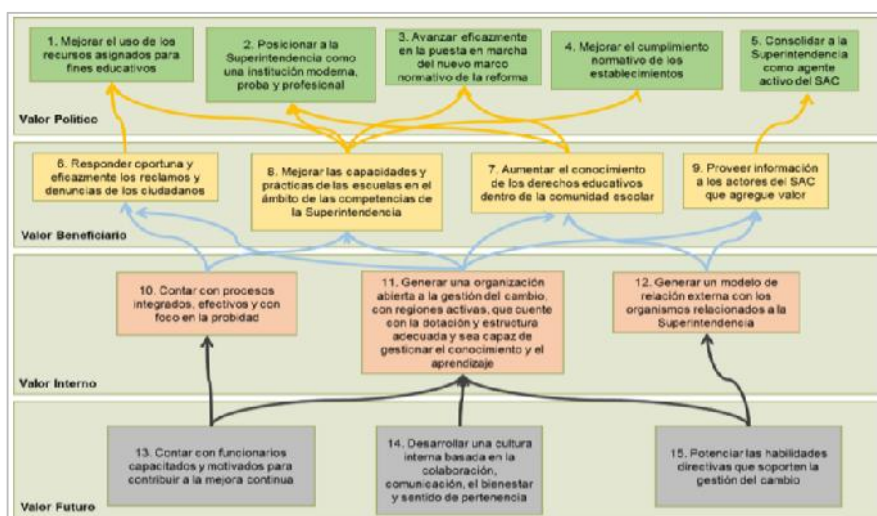


Figura N° 7: Cuadro de mando de la SUPEREDUC (SUPEREDUC, 2015)

La SUPEREDUC define cinco ejes estratégicos<sup>10</sup>: implementación de nuevas leyes (reforma educacional), nuevo modelo de fiscalización, ciudadanía y comunidad escolar, desarrollo organizacional y sistema de aseguramiento de la calidad. Dado lo anterior, el proyecto es parte del nuevo modelo de fiscalización y apoya la materialización de los siguientes objetivos e indicadores:

- Objetivo en valor político: mejorar el uso de los recursos asignados para fines educativos. Indicador: % de recursos del sistema rendidos y gastados según normativa pertinente del total de escuelas y sostenedores fiscalizados.

<sup>10</sup> Detalles disponibles en [www.supereduc.cl](http://www.supereduc.cl)

- Objetivo en valor Beneficiario: proveer información a los actores del SAC que agregue valor. Indicador: set de indicadores de cumplimiento normativo y rendición de cuentas definidos.

### **3.3 MODELO DE NEGOCIOS**

Para describir el modelo de negocios de la SUPEREDUC se utiliza el modelo Canvas desarrollado por Osterwalder & Pigneur (2009).

#### **3.3.1 Clientes**

Los clientes principales de la SUPEREDUC son los ciudadanos, segmentados en: estudiantes, en los niveles de educación parvularia, básica y media; sus padres, apoderados o tutores; los docentes, asistentes de la educación, educadores y técnicos de párvulos; los directores y sostenedores de las escuelas; los ciudadanos comunes, que no pertenecen a los grupos anteriores; instituciones como: consejos escolares, centros de padres, instituciones del SAC, agrupaciones (sostenedores, docentes, otras), Organismos del Estado, centros de investigación, fundaciones de educación, entre otras.

#### **3.3.2 Propuesta de valor**

La SUPEREDUC para cada segmento de clientes, define una propuesta de valor.

A los estudiantes, ofrece fiscalizar eficiente y eficazmente a los sostenedores y escuelas, para que existan las condiciones que les permitan recibir una educación de calidad; junto con, resguardar y dar a conocer sus derechos educacionales<sup>11</sup>.

A los padres, apoderados, tutores, docentes, asistentes de la educación, directores, educadores y técnicos de párvulos, ofrece darles a conocer sus derechos educacionales (y los derechos de los estudiantes); y publicarles información de sostenedores y escuelas, en el ámbito de las funciones de la Organización, para la toma de decisiones.

---

<sup>11</sup><https://www.supereduc.cl/wp-content/uploads/2017/11/Gu%C3%ADa-Resguardo-de-Derechos-Una-responsabilidad-compartida.pdf>



A los clientes anteriores y al ciudadano común, adicionalmente ofrece resolver oportunamente sus consultas y en caso de realizar una denuncia, ser atendida expeditamente, analizada con celeridad y tratada con el debido resguardo de identidad; y en caso de ser una vulneración de derechos o incumplimiento normativo, fiscalizar y sancionar al sostenedor y/o escuela; en ciertas ocasiones, una denuncia puede ser derivada a un servicio de mediación, que busca acuerdos entre las partes.

A los sostenedores, ofrece dar a conocer información referente a los derechos educacionales, normativa educacional vigente, circulares y dictámenes; garantizar un trato justo e igualitario, en caso enfrentarse a un proceso administrativo sancionatorio; capacitar y resolver consultas para cada proceso declarativo y/o rendición de cuentas.

A las instituciones, ofrece información, datos estructurados y documentación; resultado de funciones y procesos de la Organización, de apoyo a la toma de decisiones.

### **3.3.3 Canales**

La SUPUREDOC utiliza los siguientes canales para difundir, acceder y evaluar sus servicios y propuestas de valor.

Canales propios y bidireccionales, equipos regionales de atención presencial de consultas y denuncias; sistema integral de atención ciudadana (SIAC), con canales de atención telefónica (+56232431000) y Web para sostenedores (<https://ptf.supereduc.cl>) y ciudadanos (<https://www.supereduc.cl/atencion-ciudadana/>), actualmente se trabaja en unificarlos; y redes sociales (Facebook: supereducCL y Twitter: supereduc\_cl).

Canales propios y unidireccionales, sitio Web institucional (<https://www.supereduc.cl>); canal de youtube (superintendencia de educación); y newsletter (envío por @mail).

Canales de socios e indirectos, avisos en medios (prensa escrita, televisivo y radial); banner en sitios Web instituciones relacionadas (Instituciones del SAC, JUNJI, Conycit, Fundación Chile, Consejo de la Infancia, Integra, Educar Chile, etc.); y canal de atención Web dispuesto para cumplir la Ley de Transparencia ([www.portaltransparencia.cl](http://www.portaltransparencia.cl)).

### **3.3.4 Relación con los Clientes**

La SUPEREDUC se relaciona con sus clientes de manera personalizada y automatizada y basa sus relaciones en fundamentos como ser reconocida y distinguida por sus clientes y ser efectiva (eficiente y eficaz) en la prestación de servicios.

Relación personalizada, al atender efectivamente consultas y denuncias; mediar conflictos de la comunidad escolar; dictar seminarios y talleres de derechos educacionales; capacitar a sostenedores en cada proceso declarativo exigido.

Autoatención, buscar preguntas frecuentes a consultas ciudadanas, revisar estado de consultas y denuncias realizadas; buscar derechos educacionales, normativa educacional vigente, circulares y dictámenes; consultar información de sostenedores y escuelas; y revisar recursos digitales de procesos declarativos y rendiciones de cuentas.

### **3.3.5 Fuentes de Ingresos**

Los ingresos de la SUPEREDUC provienen de la Ley de Reajuste del Sector Público.

### **3.3.6 Recursos Claves**

Los recursos claves que permiten a la SUPEREDUC ofrecer las propuestas de valor, llegar a los segmentos de clientes y establecer relaciones con los mismos, son:

Recursos humanos especializados, en educación, resguardo de derechos, normativas educacionales, uso de recursos financieros y fiscalización.

Recursos físicos, oficinas regionales (una por región) y nivel central (en región Metropolitana) y una flota de vehículos para apoyar la fiscalización en terreno.

Infraestructura tecnológica, de apoyo a los procesos de negocio (Web institucional, buscadores de Información de sostenedores, escuelas y normas, sistema de atención integral ciudadana, sistema de agendamiento de fiscalización, sistema de procesos administrativos sancionatorios, y sistemas de declaración y rendiciones de cuentas.

### 3.3.7 Actividades Claves

Las actividades claves de la SUPEREDUC se desprenden de la Figura N° 1 (descrita en la Sección 1.2.3 Servicios Principales), específicamente de su tercer nivel; por ejemplo, la planificación anual de fiscalización, es parte del tercer nivel en Fiscalización.

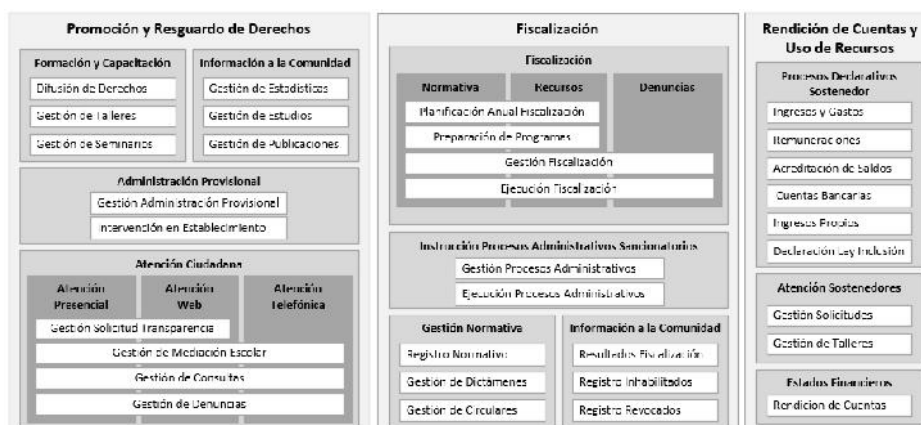


Figura N° 1: Capacidades de negocio de la SUPEREDUC

### 3.3.8 Asociaciones Claves

La SUPEREDUC no tiene alianzas estratégicas con empresas competidoras, es un monopolio de facto; tampoco tiene *joint ventures* para crear nuevos negocios o relaciones cliente - proveedor para garantizar suministros. Solo se asocia con instituciones para compartir información, en su mayoría para los procesos de fiscalización y promoción y resguardo de derechos; principales colaboradores: instituciones del SAC, Registro Civil, Servicio de Impuestos Internos, fundaciones relativas a inclusión y diversidad, agrupaciones de sostenedores (particulares subvencionados, religiosos, otros).

### 3.3.9 Estructura de Costos

La SUPEREDUC presenta una estructura de costos fijos y variables. El presupuesto<sup>12</sup> en un 70% corresponden a remuneraciones (costos fijos) y un 20% a adquisición de bienes y servicios (costos variables); el 10% restante, corresponde a inversiones.

<sup>12</sup> [http://www.dipres.gob.cl/572/articles-149470\\_Ley\\_de\\_Presupuestos\\_2017\\_V2.pdf](http://www.dipres.gob.cl/572/articles-149470_Ley_de_Presupuestos_2017_V2.pdf)

## CAPÍTULO 4: ANÁLISIS SITUACIÓN ACTUAL

### 4.1 ARQUITECTURA DE PROCESOS

Se utilizan las directrices y trabajos realizados por Barros & Julio (2011) y Barros (2015) para diagramar la arquitectura de procesos de la SUPEREDUC, ilustrada en la Figura N° 8, que cuenta con tres cadenas de valor que comparten la planificación del negocio, el desarrollo de nuevas capacidades y los procesos de apoyo. Este arquetipo según Ross, Weill & Robertson (2006) se denomina coordinación y replicación.

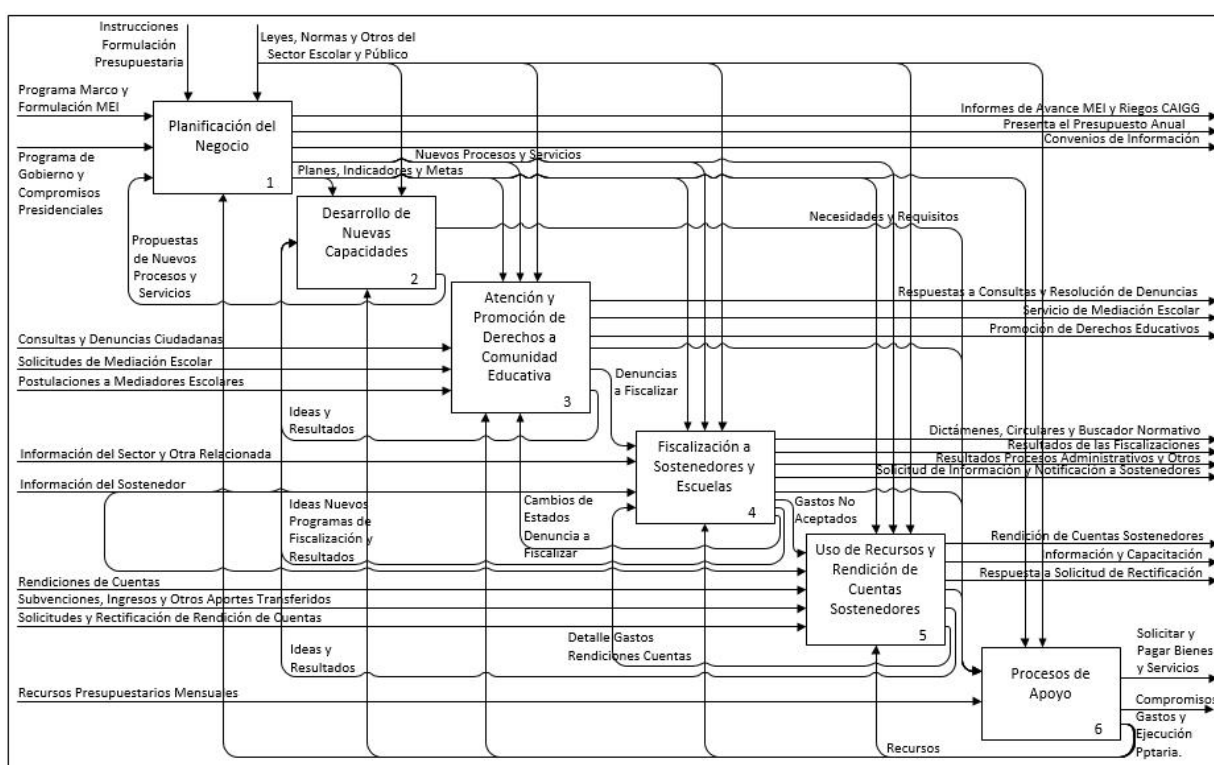


Figura N° 8: Arquitectura de procesos de la SUPEREDUC

A continuación, se describen los funcionamientos de los macroprocesos compartidos:

Planificación del negocio, la SUPEREDUC define sus planes (indicadores y metas) y nuevos procesos y/o servicios en jornadas de trabajo con directivos regionales y de nivel central. Esto en concordancia con las leyes y normas de sector escolar y público, el programa de Gobierno y compromisos presidenciales.

Este macroproceso adicionalmente, acuerda con DIPRES los indicadores y metas de eficiencia institucional<sup>13</sup> (MEI) a cumplir, define con el Consejo de Auditoría Interna General de Gobierno (CAIGG) los riesgos institucionales a prevenir y/o mitigar, tramita convenios de información con instituciones de Gobierno y presenta al Congreso Nacional el presupuesto anual para obtener su aprobación.

Desarrollo de nuevas capacidades, en la actualidad existe, pero no se encuentra estructurado; siendo instanciado, por lo general, reactivamente al entrar en vigencia una nueva ley, como fue el caso de la Ley crea la Subsecretaria de Educación Parvularia, la Intendencia de Educación Parvularia N° 20.835, 2015.

Este macroproceso independiente del actual nivel de madurez, es responsable de la innovación y del desarrollo de las propuestas de nuevos procesos y servicios a implementar; para ello, realiza estudios de perfil, pre-factibilidad y factibilidad a las ideas provenientes de las distintas cadenas de valor. Al igual que todos los macroprocesos, es normado por algunas de las más de 150 leyes y normas del sector escolar y público, por los planes (indicadores y metas) y tras enviar solicitudes de sus necesidades y requisitos a los procesos de apoyo, es provisto de los recursos para cumplir sus funciones.

Procesos de apoyo (recursos humanos, administración y finanzas y servicios tecnológicos), actualmente radican en la División de Administración General, la cual provee los recursos a los macroprocesos tras recibir las solicitudes de sus necesidades y requisitos, administra los recursos presupuestarios mensuales, paga los compromisos adquiridos a proveedores e informa mensualmente en el sistema de gestión financiero del Estado (SIGFE) los compromisos, gastos y ejecución presupuestaria.

A continuación, se describen las tres cadenas de valor, profundizando en aquellos flujos de información relevantes y relacionados a los propósitos del proyecto.

Atención y promoción de derechos a la comunidad educativa, responsable de administrar la relación con la ciudadanía a través de los canales presencial, telefónico y Web; por los

---

<sup>13</sup> Detalles de la Formulación de Metas de Eficiencia Institucional 2017, en la siguiente URL: [http://transparencia.supereduc.cl/IRIS\\_FILES/\\_T.ACTIVA/3025/formulacion\\_mei\\_2017.pdf](http://transparencia.supereduc.cl/IRIS_FILES/_T.ACTIVA/3025/formulacion_mei_2017.pdf)

cuales atiende más de 7.000 consultas y 12.000 denuncias por año. Solicitudes que provienen principalmente de los padres, apoderados o tutores (71% de las consultas y 78% de las denuncias).

Las denuncias ciudadanas dependiendo de la gravedad de los hechos pueden ser derivadas a los procesos de fiscalización, en la actualidad un 15% de los casos son investigados en terreno. En determinados casos la SUPEREDUC ofrece un servicio de mediación en busca acuerdos entre los actores de la comunidad escolar en conflicto.

Este macroproceso, desde el segundo trimestre del año 2016 realiza la promoción de los derechos educativos de los distintos actores de la comunidad escolar; utilizando para ello, boletines digitales, convenios de colaboración con fundaciones relacionadas, publicaciones en portal Web institucional, prensa escrita, capacitaciones y seminarios.

Uso de recursos y rendición de cuentas sostenedores, responsable de resguardar la legalidad del uso de los recursos transferidos por el Estado a sostenedores de establecimientos educacionales. Para ello, gestiona el proceso de declaración anual de ingresos y gastos (rendición de cuentas) y con la información obtenida, solicita al proceso de fiscalización detectar documentos riesgosos, preparar la muestra a fiscalizar, ejecutar las visitas a terreno y ser informada de aquellos gastos que no fueron aceptados - aquellos egresos detectados en una fiscalización que habiendo sido previamente objetados, la SUPEREDUC considera que no se ajustan a los fines educativos, generales o especiales, según sea el caso, o que sus respaldos documentales están adulterados o no dieron cumplimiento a los requisitos legales – y que deben ser incorporados como deuda en el próximo proceso de declaración anual de ingresos y gastos.

Este macroproceso, adicionalmente, publica información en el sitio Web institucional, ejecuta jornadas de capacitación a sostenedores a lo largo del país, y en la actualidad según el artículo 5 de la Ley N° 21.006 debe brindar el servicio de rectificación de ingresos y gastos declarados en rendiciones de cuentas de años anteriores.

Fiscalización a sostenedores y escuelas, macroproceso donde radica el proyecto y es descrito en detalle en las siguientes secciones.

## 4.2 MODELAMIENTO DETALLADO DE PROCESOS

### 4.2.1 Modelamiento IDEFO

Para instanciar los patrones de procesos para la fiscalización a sostenedores y escuelas, objeto del Proyecto, se utilizan las directrices y trabajos realizados por Barros & Julio (2011), Barros (2015), Barros (2004) y Araya (2012).

#### 4.2.1.1 Fiscalización a Sostenedores y Escuelas

La arquitectura ilustrada en la Figura N° 9 establece como los procesos y flujos de información interactúan para prevenir y/o mitigar vulneraciones de derechos educacionales, evaluar el grado de cumplimiento normativo de sostenedores y escuelas, detectar todos aquellos gastos no aceptados y pronunciarse respecto de los incumplimientos detectados (amonestaciones, sanciones y multas).

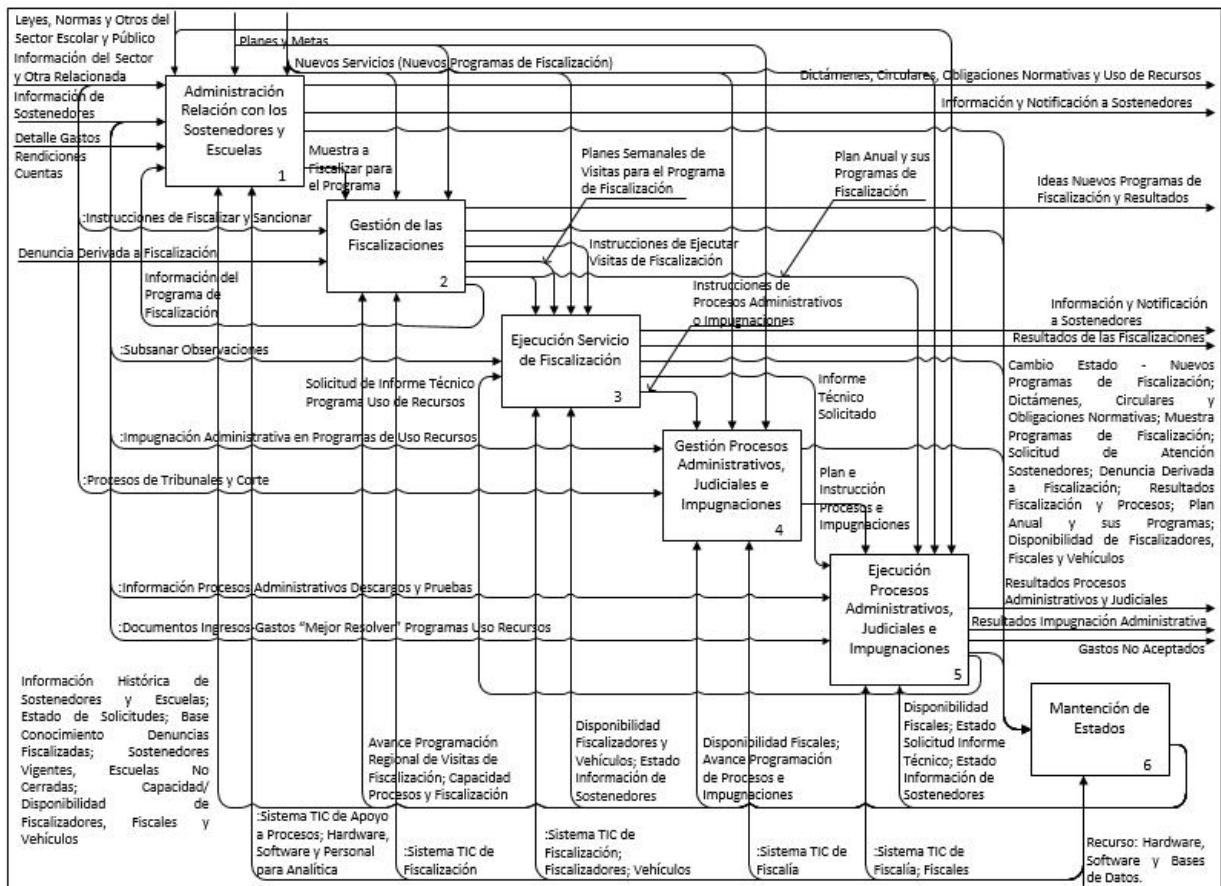


Figura N° 9: Fiscalización a sostenedores y escuelas

El funcionamiento de esta arquitectura, para los propósitos del proyecto, inicia en el macroproceso planificación del negocio al definir el plan anual de fiscalización, siendo este plan calendarizado en el proceso de gestión de las fiscalizaciones. Este último, por cada programa de fiscalización solicita al proceso administración relación con los sostenedores y escuelas preparar la muestra a fiscalizar.

La muestra a fiscalizar en el proceso de gestión de las fiscalizaciones se transforma en planes semanales, los cuales se ingresan en el sistema de fiscalización (SIFE) y donde, a cada instrucción de fiscalizar, se asocia el sujeto de fiscalización, la fecha de visita y el o los fiscalizadores.

El proceso de ejecución del servicio de fiscalización capacita a los coordinadores regionales y fiscalizadores en cada programa, de forma tal que al recibir las instrucciones de fiscalización estas sean conocidas y se cumplan los protocolos establecidos. En particular, para el programa uso de recursos SEP, objeto del proyecto, los fiscalizadores, por cada visita agendada descargan en sus laptops las hojas de trabajo (conjunto de documentos riesgosos) y la plantilla del acta de fiscalización; en el acta se registran los incumplimientos y dependiendo de la gravedad de estos, se da plazo al sostenedor o escuela para subsanar las observaciones o notificar el inicio proceso de impugnación.

Finalmente, el proceso gestión y ejecución de procesos administrativos, judiciales e impugnaciones, al no estar estructurados se resumen a asignar un fiscal a cada caso para resolver si el sostenedor o escuela es sobreseído, sancionado o multado. En el caso, de los programas de uso de recursos los fiscales deciden la legalidad de los gastos objetados (documentos riesgosos fiscalizados).

#### ***4.2.1.1.1 Administración Relación con los Sostenedores y Escuelas***

Esta arquitectura recibe del proceso de gestión de las fiscalizaciones la información del programa al cual preparar la muestra a fiscalizar y para ello, instancia el proceso de análisis de sostenedores - escuelas (ver Figura N° 10). Este último, define los criterios de riesgos y los valoriza para preparar la muestra a fiscalizar para el programa; donde, cada criterio es valorizado tras procesar la información del sector y otra relacionada, la información histórica de sostenedores y escuelas, y para los programas de uso de recursos procesa los detalles de gastos de las rendiciones de cuentas.



Esta arquitectura, adicionalmente resuelve la atención regional de sostenedores y/o personal tras acogerse a la Ley que regula el lobby y las gestiones que representen intereses particulares ante autoridades y funcionarios, N° 20.730, 2014.

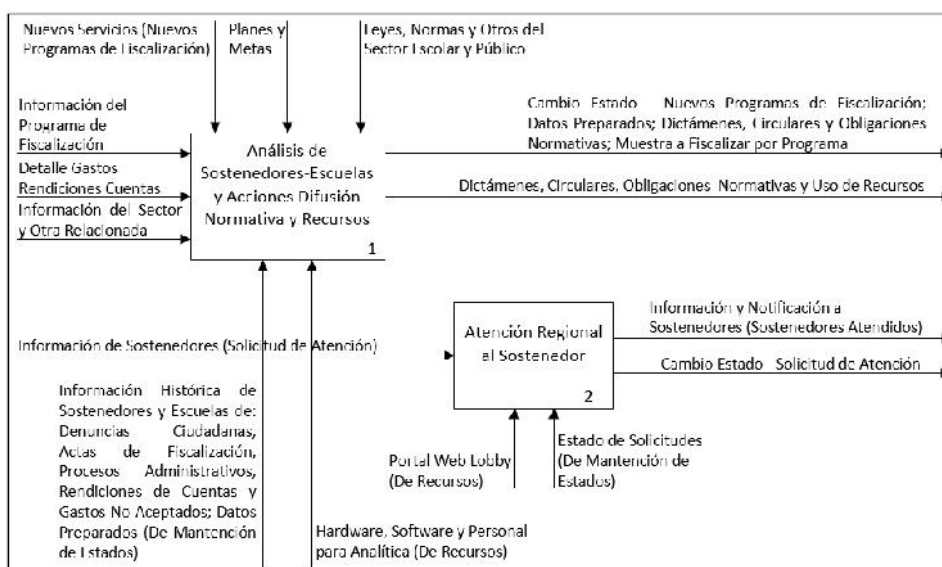


Figura N° 10: Administración relación con los sostenedores y escuelas

#### 4.2.1.1.1 Análisis de Sostenedores - Escuelas y Acciones Difusión Normativa y Recursos

Esta arquitectura en el proceso analizar comportamientos de sostenedores y escuelas (ver Figura N° 11) define los criterios de riesgos para el programa y tras procesar diversas fuentes de información, los valoriza. Cada criterio valorizado se pondera por un factor<sup>14</sup> (valor en rango 0 a 1) y con lo cual, se obtiene el ranking del programa para cada sostenedor y/o escuela.

Este ranking, junto a otras características de los sostenedores y/o escuelas, en el proceso planificar programas de fiscalización definen la muestra a fiscalizar, tras ejecutar una serie de procedimientos coordinados entre el nivel central y regional.

Esta arquitectura, adicionalmente resuelve los nuevos programas mandatados por la planificación del negocio en el proceso de introducción de nuevos programas de

<sup>14</sup> La suma de todos los factores debe sumar 1.

fiscalización y prepara, actualiza y difunde dictámenes, circulares y obligaciones normativas y/o de recursos en el proceso definir acciones de difusión normativa y recursos a cumplir por los sostenedores y escuelas.

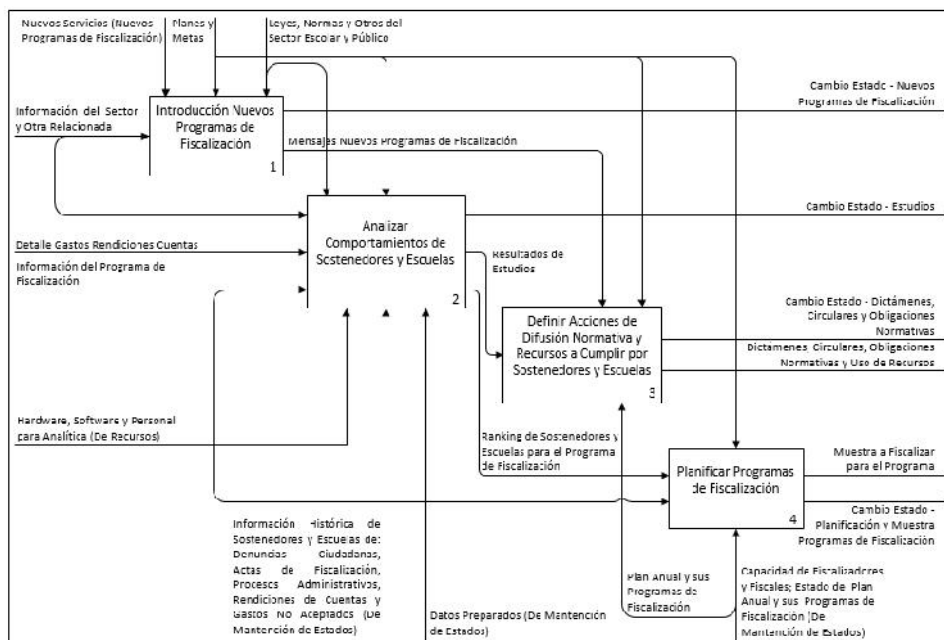


Figura N° 11: Análisis de sostenedores - escuelas y acciones de difusión normativa y recursos

#### 4.2.1.1.2 Gestión de las Fiscalizaciones

Esta arquitectura (ver Figura N° 12) establece como los procesos, actividades y flujos de información interactúan para calendarizar cada programa de fiscalización instruido por el plan anual, definir los planes semanales de visitas de fiscalización, agendar las visitas de fiscalización y efectuar el seguimiento y control del plan anual y sus programas.

El proceso de planificación anual de fiscalización y sus programas es quien crea en el SIFE cada programa de fiscalización, asocia las materias y procedimientos a aplicar; junto con, registrar la fecha de inicio y fin y la cantidad de visitas a efectuar.

El proceso de planificación regional de fiscalización se reciben las denuncias a investigar en terreno derivadas por el macroproceso atención y promoción de derechos a comunidad educativa, las muestras a fiscalizar de cada programa y las instrucciones de fiscalizar provenientes de otras instituciones (CGR, MINEDUC, Agencia de la Calidad de

la Educación u otras). Estos en cada región, se priorizan y se transforman en planes semanales de visitas de fiscalización. Cada plan, en el proceso decidir asignaciones y agendamientos para visitas de fiscalización es ingresado al SIFE por el encargado regional y quien, a cada visita asigna uno o más fiscalizadores y una fecha de ejecución.

Los procesos de control en la ejecución de los planes manualmente preparan informes de seguimientos semanales que dan cuenta de un set de indicadores agregados y donde, ningún indicador alerta de retrasos en la ejecución de los servicios de fiscalización versus lo planificados o si bien estos servicios fueron ejecutados, no alertan de retrasos en el registro del trabajo en el SIFE; impactando con ello, próximas planificaciones.

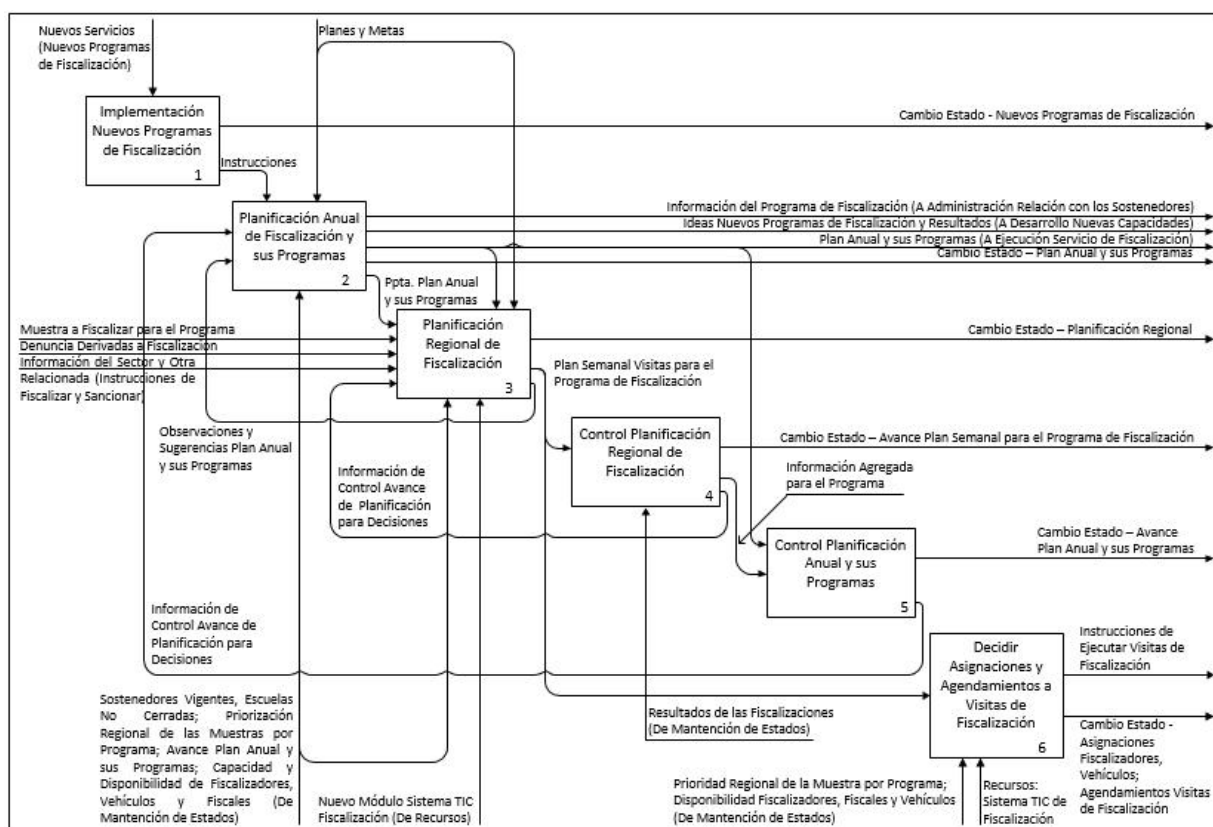


Figura N° 12: Gestión de las fiscalizaciones

#### 4.2.2 Modelamiento BPMN

En esta sección se ilustran las prácticas de trabajo de interés para los propósitos del proyecto, las cuales son: preparar ranking de sostenedores y/o escuelas, alojada en el proceso analizar comportamientos de sostenedores y escuelas; y obtener la muestra a

fiscalizar, alojada en el proceso planificar programas de fiscalización. Ambas instanciadas para el programa de fiscalización de uso de recursos SEP.

#### 4.2.2.1 Preparar Ranking de Sostenedores y/o Escuelas

Esta práctica de trabajo inicia en el evento que provee la información del programa de fiscalización, proveniente del proceso de gestión de las fiscalizaciones y dirigido al jefe unidad de datos, tal como ilustra la Figura N° 13. Con esta información, el jefe unidad de datos designa un analista responsable de preparar el ranking de sostenedores y/o escuelas para idear los criterios de riesgo y validarlos con el responsable del programa y con el analista de datos asignado (por el jefe unidad de datos).

El analista con los criterios de riesgo consensuados prepara un informe para revisión del jefe unidad de datos, quien puede visar u observar cada criterio. Cada criterio de riesgo aprobado y que no haya sido utilizado en años previos, el analista prepara las ecuaciones, consultas a base de datos (SQL) o algoritmos que permitan valorizarlo; en paralelo los analistas de datos preparan la base analítica asociada a cada criterio.

El analista, tras ejecutar los criterios de riesgo analiza los resultados y reprocessa hasta obtener conformidad; luego, reúne los resultados (listado de documentos riesgosos) y los pondera por su factor (valor en rango 0 a 1 y donde la suma de todos los factores es 1) de cada criterio y calcula el ranking de sostenedores y/o escuelas para el programa. Este informe es visado por el jefe unidad de datos y es enviado al proceso planificar programas de fiscalización.

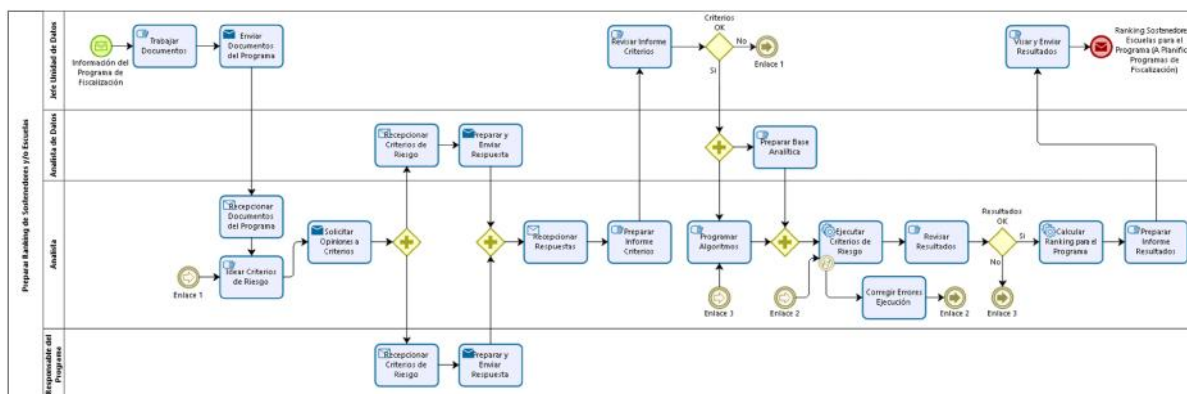


Figura N° 13: Preparar ranking de sostenedores y/o escuelas (as is)

#### 4.2.2.2 Obtener la Muestra a Fiscalizar

Esta práctica de trabajo inicia con el evento ranking sostenedores y/o escuelas para el programa, proveniente del proceso analizar comportamientos de sostenedores y escuelas dirigido al responsable del programa, tal como ilustra la Figura N° 14. Con esta información el responsable del programa prepara la propuesta de muestra que cumpla las siguientes restricciones: tamaño de la muestra a fiscalizar, tamaño de la muestra regional, cantidad definida de sostenedores y/o escuelas municipales, cantidad definida de sostenedores y/o escuelas particulares subvencionadas, cantidad definida de sostenedores y/o escuelas nunca fiscalizadas y conjunto fijo de sostenedores y/o escuelas. La propuesta de muestra es enviada al jefe fiscalización quien decide si visar o reprocesar con valores diferentes en algunos de los parámetros (restricciones).

La propuesta de muestra visada por el jefe fiscalización es enviada al jefe de división de fiscalización quien decide si visar o reprocesar con valores diferentes en algunos de los parámetros (restricciones) repitiendo el circuito. La propuesta definitiva es enviada a los jefes regionales fiscalización y quienes, dentro de un plazo pueden solicitar agregar o quitar elementos a la propuesta de muestra justificando cada cambio. Revisadas las solicitudes regionales, el jefe de división de fiscalización define la muestra a fiscalizar para el programa y es enviada al proceso de gestión de las fiscalizaciones.

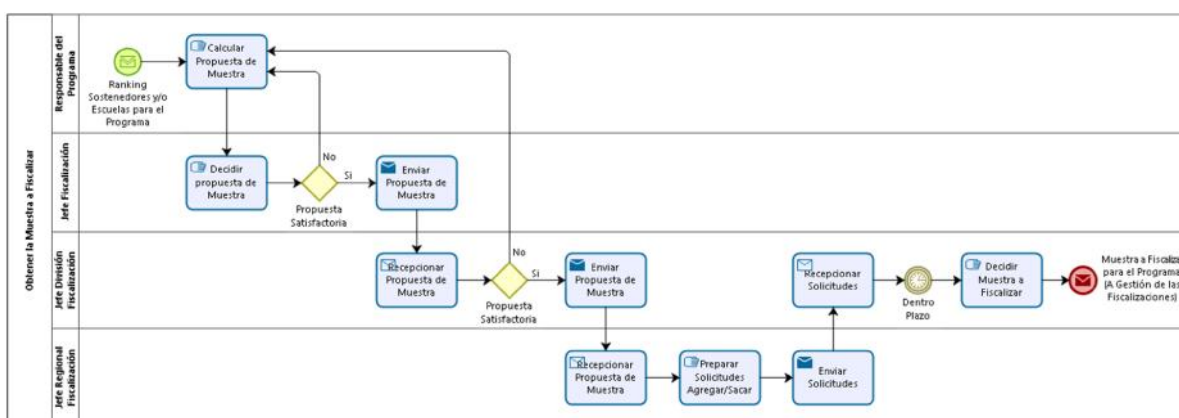


Figura N° 14: Obtener la muestra a Fiscalizar (as is)

### 4.3 DIAGNÓSTICO DE LA SITUACIÓN ACTUAL

En la práctica de trabajo preparar ranking de sostenedores y/o escuelas, perteneciente al proceso analizar comportamientos de sostenedores y escuelas (descrita en la sección 4.2.2.1), los criterios de riesgo utilizados para calcular el ranking de sostenedores y escuelas para programa uso de recursos SEP en gran mayoría, identifican documentos riesgosos usando reglas de control que utilizan valores explícitos en los atributos<sup>15</sup> de los registros de gastos declarados por los sostenedores en la rendición de cuentas SEP; ejemplos de estos son:

- Rut proveedor inválido: identifica si el RUT de proveedor del documento de gasto emitido (boleta, factura, otro) no es válido por Servicio Impuestos Internos.
- Proveedor ATE inválido: identifica si el RUT de proveedor del documento de gasto imputado en la cuenta<sup>16</sup> asesorías técnicas pedagógicas (ATE) no está reconocido y/o autorizado por el MINEDUC para prestar este tipo de servicios.
- Rango fechas: identifica si el documento de gasto de un producto adquirido o servicio contratado no cumple con el rango de fechas definido.
- Gasto Alto: identifica todos los documentos de gastos imputados en una cuenta específica y donde, la suma de montos supera un determinado valor (parámetro).

Conocido lo anterior, los actuales criterios de riesgo, principalmente no responden a:

- Reflejar si el comportamiento del sostenedor y/o escuela tiene mayor probabilidad de incumplir el debido uso de recursos SEP y con ello, focalizar los recursos de la SUPEREDUC en aquellos sostenedores y/o escuelas que presenten mayores niveles de riesgo.

---

<sup>15</sup> Según SUPEREDUC (2016) un registro de documento de gasto rendido es la tupla de atributos: RBD/AC; subvención; código de cuenta; tipo de documento; número documento; fecha documento; fecha pago; descripción del gasto; RUT proveedor; nombre proveedor; monto del gasto; monto del documento.

<sup>16</sup> Detalles de las cuentas de gastos y sus reglas <https://ptf.supereduc.cl/manuales.html>

- En particular para la SEP, al ser estos recursos de uso exclusivo y que van en directo apoyo a los estudiantes más vulnerables del país, los actuales criterios de riesgo escasamente clasifican los documentos de gastos rendidos por los sostenedores según el uso dado, descrito y declarado, en la rendición de cuentas.
- La mayoría de los criterios de riesgo deben ser restricciones a implementar en el aplicativo tecnológico dispuesto a los sostenedores para realizar la rendición de cuentas; dicho eso, es necesario explorar nuevos y novedosos criterios de riesgos.

En el proceso analizar comportamientos de sostenedores y escuelas, la práctica de trabajo preparar ranking de sostenedores y/o escuelas (descrita en la sección 4.2.2.1) representa casi la total función del proceso. En particular, para los programas de uso de recursos que fiscalizan la legalidad de los gastos y verifican la existencia de los bienes comprados y/o la prestación de los servicios contratados por los sostenedores en fecha posterior a la declaración anual de gastos, este proceso no cuenta con prácticas de trabajo que provean información durante el año que permita tomar medidas para evitar y/o mitigar riesgos derivados de usos indebidos de recursos.

En la práctica de trabajo obtener la muestra a fiscalizar, perteneciente al proceso planificar programas de fiscalización (descrita en la sección 4.2.2.2), la tarea que transforma el ranking de sostenedores y/o escuelas en la propuesta de muestra a fiscalizar para el programa, no es tarea simple y objetiva, además de no contar con apoyo tecnológico (software apropiado) y un algoritmo de programación matemática (optimización) que permita setear un conjunto de parámetros y así, determinar la asignación óptima de escuelas y/o sostenedores que maximicen el beneficio (en el caso de los programas de uso de recursos, maximice el reintegro de recursos). De igual manera, el flujo de trabajo e interacciones con los distintos roles, es manual; pudiendo ser apoyado con tecnológica para gestionar, trazar y almacenar información resultante.

En el proceso de gestión de las fiscalizaciones (descrito en la sección 4.2.1.1.2), si bien el proyecto no radica en este proceso, es necesario intervenirlo para cumplir uno de sus objetivos; por ello, se debe dar visibilidad online y sistematizada (*dashboards*) a las jefaturas y fiscalizadores respecto de los cambios de estado de los agendamientos planificados versus los servicios ejecutados, ambos registrados en el SIFE. Este cambio

de práctica apunta a: mejorar la calidad en la información de disponibilidad de fiscalizadores, elaborar un predictor (basado en inferencia estadística) de duración para cada tipo de fiscalización y aumentar la capacidad de visitas de fiscalización al efectuar seguimiento y control con una herramienta online y sistematizada.

En el proceso de ejecución del servicio de fiscalización (descrito en la sección 4.2.1.1), si bien el proyecto no radica en este proceso, se infiere que no cuenta con un algoritmo y apoyo tecnológico que permita ejecutar el servicio de fiscalización en modalidad digital/virtual, que en cierto grado no requiera de visitas a terreno, optimizando el proceso y disminuyendo burocracias.

#### **4.4 CUANTIFICACIÓN DEL PROBLEMA U OPORTUNIDAD**

En el año 2016 de la totalidad de fondos SEP transferidos por el Estado, la SUPEREDUC identificó la suma de MM\$ 76.000 en documentos riesgosos (representa un 9,5% del total de fondos SEP) que, tras ejecutar 2.100 visitas<sup>17</sup> de fiscalización reintegraron sólo MM\$ 22.250 al sistema escolar. Con estas cifras se desprende que en promedio una visita a terreno reintegra MM\$ 10,5 y la efectividad del programa alcanza un 30%.

- El rendimiento es menor si se compara la SUPEREDUC con otras instituciones que realizan labores similares, que detectan tasas de evasión<sup>18</sup> sobre 20% y con países europeos, que declaran tener tasas menores a 10%. Con este antecedente, cada punto porcentual adicional logrado con los nuevos y novedosos criterios de riesgo para el programa, corresponde a MM\$ 8.000.
- Los resultados son similares al comparar el programa uso de recursos SEP 2016 con el año anterior. La tasa de reintegro se mantiene en un 3%. Cada punto porcentual adicional en la tasa, que se logre con los nuevos y novedosos criterios de riesgo para el programa, aumenta el reintegro en MM\$ 767.

---

<sup>17</sup> Fuente: Unidad de Datos de la SUPEREDUC.

<sup>18</sup> Fuentes: <http://www.uss.cl/economia-y-negocios/disminuyo-la-evasion-del-iva-chile-gracias-nuevo-modelo-fiscalizacion/> y [http://www.sii.cl/aprenda\\_sobre\\_impuestos/estudios/Antecede.htm](http://www.sii.cl/aprenda_sobre_impuestos/estudios/Antecede.htm)



## **CAPÍTULO 5: PROPUESTA DE DISEÑO DE PROCESOS**

### **5.1 DIRECCIONES DE CAMBIO Y ALCANCE**

Se utiliza el trabajo realizado por Barros (2004) para establecer la dirección del cambio del proyecto, que debe estar relacionado con el modelo de negocio y el patrón que sirve de guía; además, de explicitar las variables de rediseño y orden para ser consideradas.

#### ***5.1.1 Estructura de la Empresa y Mercado***

Esta variable no es relevante para los propósitos del proyecto, debido a que la SUPEREDUC es un monopolio de facto creado por Ley, y la cual, explicita sus funciones y obligaciones; además, la Organización no declara intenciones de moverse de su posicionamiento estratégico de canal exclusivo.

#### ***5.1.2 Asignación de Responsabilidades***

Esta variable es relevante para el proyecto al asignar responsabilidades de manera diferente al proceso analizar comportamientos de sostenedores y escuelas, que le permitan preparar fiscalizaciones preventivas y focalizadas para el programa uso de recursos SEP, incorporando analítica basada en datos; además de, generar recomendaciones oportunas a los sostenedores que se encuentran realizando la rendición de cuentas, de modo de evitar y/o mitigar usos indebidos de recursos y alertar errores de imputación de gastos.

Estas nuevas responsabilidades, desde luego, impactan en la actual estructura organizacional de la SUPEREDUC, al pasar de una configuración planificada e instanciada en fecha posterior a la declaración anual de gastos (en caso del programa uso de recursos SEP) a una estructura cada vez más preventiva y educativa, y en un futuro cercano, cada vez más digital al no depender exclusivamente de la capacidad de visitas a terreno para el cumplimiento de metas y objetivos.

#### ***5.1.3 Anticipación***

La idea principal de esta variable es analizar las acciones para anticiparse a eventos futuros, a través del desarrollo de una capacidad de predicción de tales eventos; por ello,

esta variable es relevante para los propósitos del proyecto, al ser uno de sus objetivos el segmentar y clasificar a sostenedores y escuelas en función de sus características y comportamientos relacionados con la gestión financiera, de modo que se tomen acciones preventivas y focalizadas para disminuir usos indebidos de recursos SEP.

Adicionalmente, y tal como es descrito en el diagnóstico de la situación actual, en la medida que las reglas que rigen la mayoría de los criterios de riesgo se programen en el sistema TI utilizado por los sostenedores para rendir cuentas, la capacidad actual de detección de documentos riesgosos se reduce considerablemente. Sin embargo, esto no implica que se esté haciendo un buen uso de los recursos, por lo que es necesario anticiparse a este escenario y desarrollar e innovar en nuevos criterios de riesgo que detecten nuevas formas o nuevos tipos de documentos riesgosos que surgirán o que hoy están presentes, pero se encuentran ocultos a los actuales criterios de riesgo.

#### ***5.1.4 Integración de Procesos Conexos***

Con esta variable se determina el grado de interconexión de los procesos y como estos deben ser diseñados. En base a lo anterior, para efectuar fiscalizaciones preventivas y focalizadas en el uso de recursos SEP, los procesos planificar programas de fiscalización, analizar comportamientos de sostenedores y escuelas, gestión de las fiscalizaciones y ejecución servicio de fiscalización deben funcionar en estrecha interacción y deben ser rediseñados en conjunto, al pertenecer a la misma cadena de valor y tener relación de precedencia. Respecto a las acciones preventivas y/o mitigación relacionadas al debido uso de recursos SEP requeridas para reducir el monto actual de gastos no aceptados, los procesos planificar programas de fiscalización y analizar comportamientos de sostenedores y escuelas deben trabajar en estrecha relación y ser rediseñados en conjunto para proveer información permanente a los procesos: uso de recursos y rendición de cuentas sostenedores, el cual publicará las recomendaciones en el portal Web dispuesto para declarar gastos; y gestión de las fiscalizaciones y ejecución servicio de fiscalización, que aprovecharán cada visita a terreno para efectuar breves capacitaciones individualizadas que utilicen las recomendaciones preparadas y/o efectuar seguimiento y control de la subsanación de estas.

### **5.1.5 Mantención Consolidada de Estados**

Esta variable define como se obtienen los datos necesarios para la ejecución de los procesos rediseñados. Desde el ingreso de datos dentro de un mismo proceso hasta el uso de datos de sistemas externos a la Organización. En base a lo anterior, y dado el bajo nivel de madurez de la SUPEREDUC en esta materia, el proyecto para cumplir sus propósitos debe implementar un repositorio de información, en una versión incipiente y escalable a otros programas de fiscalización, en el cual almacenar estructuradamente la información recopilada de los procesos internos como externos a la Organización; lo anterior, es base para ejecutar analítica basada en datos que provean información útil a los procesos de fiscalizaciones preventivas. De igual manera, para trazar los cambios de estado de los flujos de información relativos a las fiscalizaciones para el programa objeto del proyecto, se debe implementar un *datamart* a los procesos de gestión de las fiscalizaciones, ejecución del servicio de fiscalización y rendición de cuentas de sostenedores, con los respectivos cubos OLAP.

### **5.1.6 Coordinación**

Esta variable declara las necesidades de coordinación organizacional que requieren los procesos a diseñar. En base a lo anterior, el proyecto utiliza reglas formales de coordinación entre las unidades responsables de los procesos: planificar programas de fiscalización, analizar comportamientos de sostenedores y escuelas, gestión de las fiscalizaciones, ejecución servicio de fiscalización y rendición de cuentas de sostenedores; las cuales, disminuyan la utilización de jerarquías para resolver situaciones frecuentes, y sólo se recurra a ellas, para resolver excepciones.

En cuanto a la definición de reglas, por ejemplo, el proyecto por medio de la solicitud de preparar un programa de fiscalización preventivo, se disponga de los modelos analíticos vigentes que segmentan y clasifican a sostenedores y escuelas con comportamientos riesgosos para los propósitos del programa en cuestión y la base analítica adecuada, y con lo cual, ejecutar los modelos e identificar aquellos sostenedores y/o escuelas con mayor probabilidad de incumplimiento, definir la muestra, efectuar el agendamiento y ser fiscalizados oportunamente según protocolos establecidos y sin requerir la excesiva utilización de jerarquías regionales y del nivel central, tal como ocurre en la actualidad.

### **5.1.7 Prácticas de Trabajo**

Esta variable materializa y detalla las opciones de diseño de las variables definidas antes. Estas prácticas deben permitir la ejecución de las actividades de los procesos rediseñados. En base a lo anterior, el proceso analizar comportamientos de sostenedores y escuelas, debe desarrollar prácticas relativas a conocer y entender los procesos de los sostenedores y escuelas para así, plantear hipótesis, a las cuales diseñar, validar y mantener modelos analíticos que permitan preparar el ranking de sostenedores y/o escuelas requeridos por el proceso de planificar programas de fiscalización; de igual forma, debe desarrollar las prácticas que permitan revisar periódicamente la información de gastos rendida por los sostenedores durante el proceso declarativo (no solo al final) para preparar recomendaciones al proceso de planificar programas de fiscalización.

El proceso de planificar programas de fiscalización debe desarrollar prácticas que permitan ejecutar los modelos analíticos e identificar aquellos sostenedores y/o escuelas con mayor probabilidad de incumplimiento, definir la muestra, efectuar el agendamiento y lograr fiscalizarlos oportunamente y sin requerir la excesiva utilización de jerarquías regionales y del nivel central, tal como ocurre en la actualidad. Respecto a las recomendaciones de uso de recursos, se debe desarrollar la práctica que permita decidir si se fiscaliza en terreno, es publicada en el portal Web de rendición de cuentas o ambas.

La Organización para pasar de una total estructura planificada a una estructura cada vez más preventiva y educativa, adicionalmente debe trabajar en las prácticas de trabajo en los procesos de gestión de las fiscalizaciones y ejecución servicio de fiscalización que guarden relación con el agendamiento y ejecución, oportunas; para ello, debe establecer nuevos y automatizados mecanismos de seguimiento y control.

### **5.1.8 Apoyo Computacional**

Esta variable es una consecuencia de las anteriores, al satisfacer cada uno de los requerimientos. Por ello, se identifican las propiedades del programa uso de recursos SEP y gestión financiera de sostenedores y escuelas; a las cuales, asociarles datos de procesos internos como externos a la Organización, y a la vez, diseñar una versión inicial y escalable de un *datawarehouse* (DW), al cual cargar los datos procesados. Además, se requiere una base analítica, resultante del procesamiento de datos con formato semi -

estructurado de las rendiciones de cuentas realizadas por los sostenedores; siendo necesario para esta labor, diseñar e implementar una herramienta tecnológica ad hoc.

Para continuar con la generación de valor a través de los datos, se incorpora analítica al proceso de analizar comportamientos de sostenedores y escuelas. Para ello, es necesario un software especializado, que en un principio permita descubrir información del comportamiento de sostenedores y escuelas en el ámbito de la gestión financiera, y desarrollar nuevos criterios de riesgo para el programa uso de recursos SEP; información útil, eficiente y focalizada, de apoyo a la toma de decisiones para los procesos de fiscalización preventivas y recomendaciones de uso de recursos a sostenedores.

En apoyo a las necesidades de integración de procesos conexos, mantención de estados consolidada y coordinación; se debe implementa un *datamart* a los procesos de gestión y ejecución de las fiscalizaciones, que utilice la información transaccional del actual sistema TIC que los soporta. Junto con, implementar un set de indicadores (cubos OLAP) que den visibilidad a tiempos ociosos o sobretiempos (foco capacidad fiscalizadora), y en una versión posterior, medir la efectividad de las fiscalizaciones preventivas efectuadas y las recomendaciones realizadas a sostenedores.

## 5.2 DISEÑO DETALLADO DE PROCESOS TO BE

### 5.2.1 Administración Relación con los Sostenedores

La arquitectura de este proceso respecto a su situación actual, modifica su funcionamiento en el proceso análisis de sostenedores - escuelas y acciones difusión normativa y recursos al incorporar los flujos de información de datos, variables, análisis y modelos *data mining*, evaluación de modelos *data mining* y modelos analíticos de *data mining* vigentes y disponibles, resaltados en la Figura N° 15.

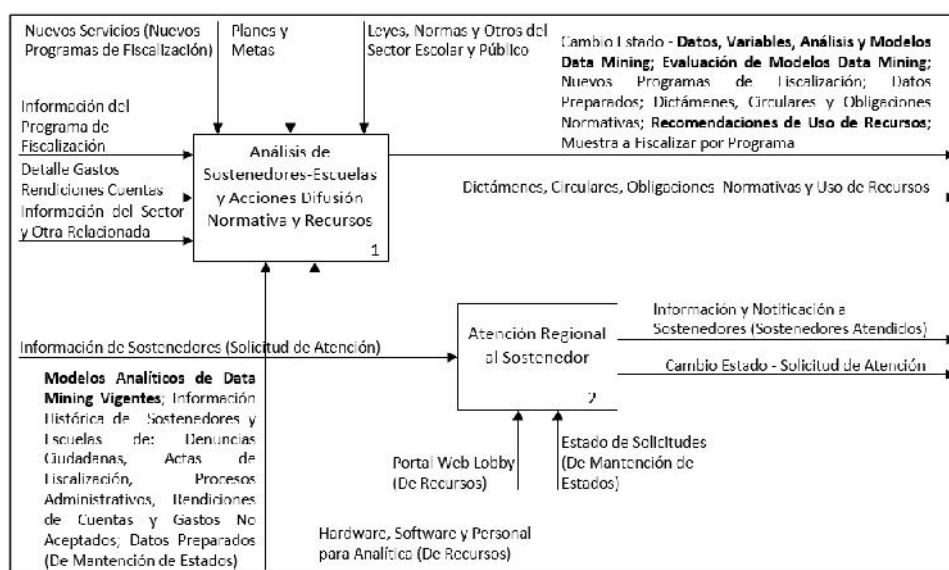


Figura N° 15: Administración relación con los sostenedores (to be)

En el proceso análisis de sostenedores - escuelas y acciones de difusión normativa y recursos, es donde se incorpora la analítica basada en datos al disponer modelos de *data mining* requeridos e instanciados por el proceso planificación del programa de fiscalización para decidir la muestra de sostenedores y/o escuelas a fiscalizar con enfoque preventivo y focalizado para el programa uso de recursos SEP. De igual manera, en este proceso se preparan recomendaciones relacionadas con la rendición de cuentas a sostenedores y que serán publicadas tras contar con visación del proceso planificación del programa de fiscalización.

Al instanciar el proceso análisis de sostenedores - escuelas y acciones de difusión normativa y recursos, producto de los nuevos flujos de información, se obtiene la arquitectura ilustrada en la Figura N° 16, que resaltan los procesos y los flujos de

información que modifican su funcionamiento respecto a su situación actual, y que darán origen a las prácticas de trabajos propuestas por el proyecto.

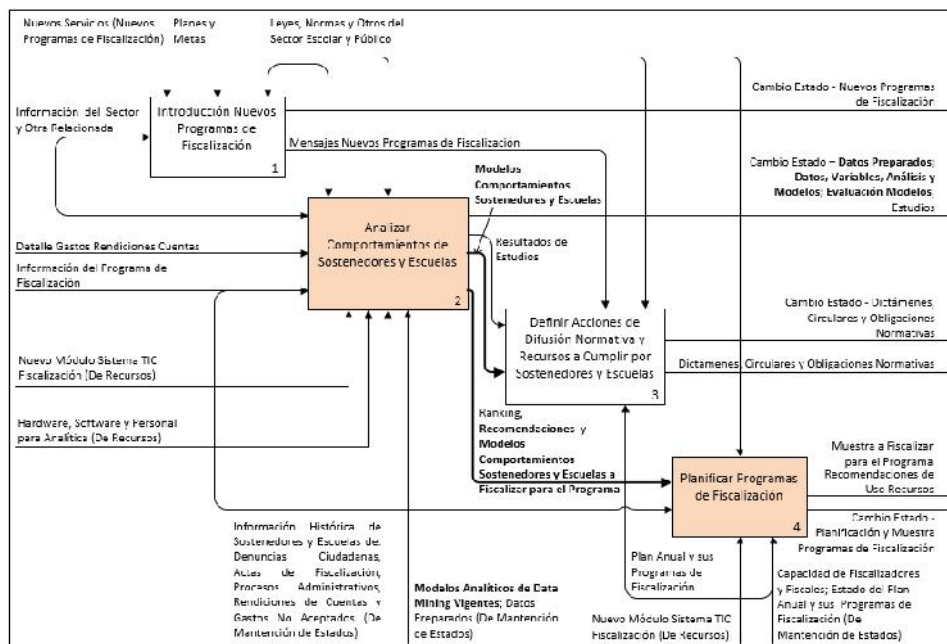


Figura N° 16: Análisis de sostenedores – escuelas y acciones difusión normativa y recursos (to be)

Al instanciar el proceso analizar comportamientos de sostenedores y escuelas se obtiene la arquitectura de la Figura N° 17, la cual es un rediseño completo respecto a su situación actual y en el cual conviven la operación tradicional y la propuesta por el proyecto para preparar fiscalizaciones preventivas y focalizadas.

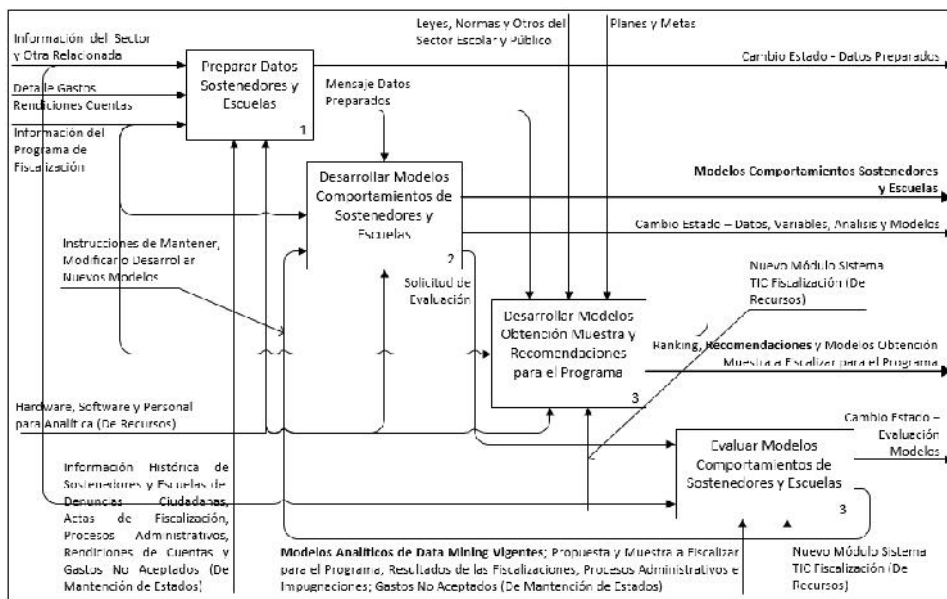


Figura N° 17: Analizar comportamientos de sostenedores y escuelas (to be)

A continuación, se describe el funcionamiento de la arquitectura de la Figura N° 17 para los propósitos del proyecto, la cual inicia con la información del programa de fiscalización uso de recursos SEP proveniente del proceso gestión de las fiscalizaciones; esta información es enviada con antelación al comienzo del proceso de rendición de cuentas e instancia la mayoría de los procesos y/o actividades de esta arquitectura.

Al ser instanciada la actividad preparar datos sostenedores y escuelas, comienza la preparación de la base de datos analítica para el programa de fiscalización uso de recursos SEP; para ello, recopila y procesa información del sector y otra relacionada, información histórica de comportamiento de sostenedores y escuelas, y los gastos de rendiciones de cuentas. Los detalles de esta actividad se presentan en la Sección 5.2.2 Diseño en BPMN.

Al ser instanciada la actividad desarrollar modelos de comportamientos de sostenedores y escuelas, comienza el diseño y validación de los modelos analíticos, y con ello el descubrimiento de patrones a ser utilizados en el proceso de planificar programas de fiscalización preventivos; para lo cual, se aplica el método crisp - dm descrito en Capítulo 2: Marco Teórico, base de la arquitectura presentada en la Figura N° 18 y cuyos detalles son descritos en la Sección 5.2.2 Diseño en BPMN.



Al ser instanciada la actividad desarrollar modelos obtención muestra y recomendaciones para el programa, comienza la preparación del ranking de sostenedores y/o escuelas para los programas de fiscalización tradicionales en estructura y forma de operar presentada en la Sección 4.2.2 Modelamiento BPMN; en el caso del programa de fiscalización uso de recursos SEP, en su enfoque tradicional, instancia esta actividad tras finalizar el proceso declarativo de rendición de cuentas. Adicionalmente, esta actividad, en su situación deseada comienza a preparar recomendaciones preventivas y/o acciones de mitigación para apoyar tempranamente a cumplir los objetivos de los programas de fiscalización; en particular, para los programas de uso de recursos, objeto del proyecto, estas recomendaciones inicialmente provienen de ejecutar periódicamente y durante todo el periodo declarativo de gastos el nuevo criterio de riesgo que clasifica las descripciones de gasto en cada documento rendido por los sostenedores (ver Figura N° 19).

Al instanciar la actividad evaluar modelos de obtención de muestras a fiscalizar por programa, comienza la toma de decisiones durante y post ejecución del programa de fiscalización respecto de: mantener, cambiar, modificar o desarrollar nuevos modelos; durante y post ejecución del programa de fiscalización; para ello, se evalúa la efectividad de la muestra fiscalizada periódicamente en función de la cantidad de aciertos logrados. Los detalles de esta actividad se presentan en la Sección 5.2.2 Diseño en BPMN.

La siguiente arquitectura establece las actividades y flujos de información resultantes de instanciar el proceso desarrollar modelos de comportamientos de sostenedores y escuelas, y tal como fue descrito se basa en el método crisp-dm. Los detalles de estas actividades se presentan en la Sección 5.2.2 Diseño en BPMN.

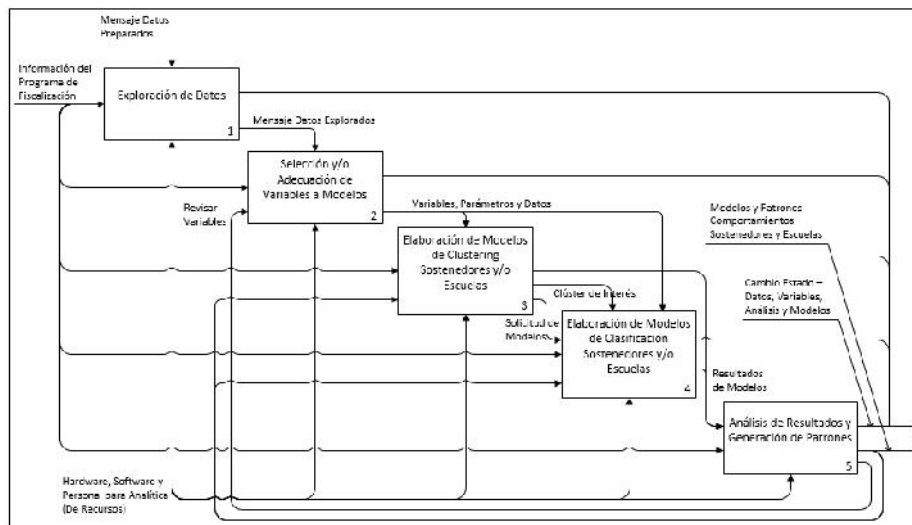


Figura N° 18: Desarrollar modelos comportamientos de sostenedores y escuelas

A continuación, se describe la arquitectura del proceso desarrollar modelos obtención muestra y recomendaciones para el programa, y tal como fue descrito sus actividades se diferencian en la oportunidad en que aplican los controles o criterios de riesgo para un programa de fiscalización determinado, ya sea para obtener el ranking de sostenedores y/o escuelas para un programa o para preparar recomendaciones que apoyen la prevención y/o mitigación de incumplimientos (ver Figura N° 19).

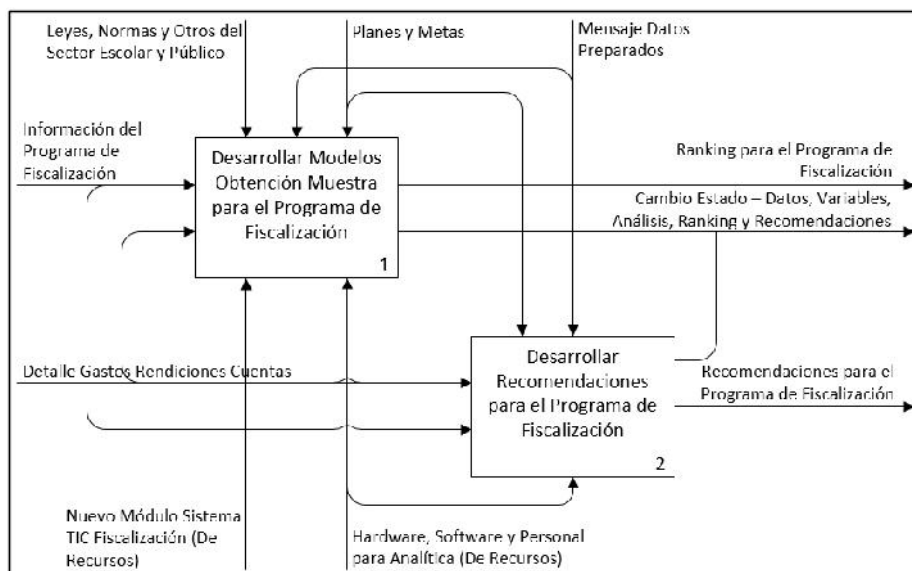


Figura N° 19: Desarrollar modelos obtención muestra y recomendaciones para el programa

La actividad desarrollar modelos de obtención muestra para el programa de fiscalización, es responsable de obtener el ranking de sostenedores y/o escuelas para los programas de fiscalización tradicionales, la cual mantiene las actividades y flujos de información presentados en Sección 4.2.2 Modelamiento BPMN.

La actividad desarrollar recomendaciones para el programa de fiscalización, es responsable de preparar información para generar acciones preventivas y/o de mitigación que apoyen a cumplir tempranamente los objetivos del programa. Esta actividad a diferencia de la anterior, descubre información al aplicar los controles o criterios de riesgo en etapas tempranas; en el caso, del programa uso de recursos SEP objeto del proyecto, los criterios de riesgo presentados en la Sección 4.3 Diagnóstico de la Situación Actual pueden ser aplicados durante todo el periodo declarativo de gastos y los documentos riesgosos identificados, serán la base para la formulación de las recomendaciones.

La relevancia adicional que encubre esta actividad, desarrollar recomendaciones para el programa de fiscalización, radica en que la SUPEREDUC no cuenta con capacidad ilimitada para efectuar visitas a terreno para fiscalizar; por ejemplo, fiscalizar documentos riesgosos en el programa uso de recursos SEP. Por lo anterior, esta actividad demostrará su validez al utilizar el nuevo criterio de riesgo (control) propuesto por el proyecto, el cual clasifica cada descripción del gasto en: uso indebido, error de imputación en cuenta de gasto o escasa descripción del gasto; las cuales, en procesos posteriores serán: publicadas en portal Web de rendición de cuentas; incorporadas como un ítem en cada visita de fiscalización preventiva a terreno para asesorar-educar al sostenedor y/o escuela, o efectuar el seguimiento y control de subsanación al ser incluidas como una materia en otros programas de fiscalización tradicionales (aprovechando la visita a terreno).

Las arquitecturas descritas para los propósitos del proyecto, en resumen, tienen por objeto generar y mantener modelos analíticos que proveen información respecto de fiscalizaciones preventivas y focalizadas para los programas de uso de recursos, y preparar recomendaciones periódicas individualizadas a sostenedores respecto del debido uso de los fondos transferidos por el Estado.

A continuación, se describen las arquitecturas que ejecutan los modelos analíticos desarrollados y dan utilidad a la información generada, al ser instruida a los procesos o herramientas de resguardo de derechos que cuenta la SUPEREDUC, la fiscalización a sostenedores y/o escuelas y la rendición de cuentas.

Para lograr lo anterior, la arquitectura del proceso planificar programas de fiscalización debe modificar su funcionamiento (ver Figura N° 20) respecto a su situación actual para incorporar actividades y flujos de información que permitan preparar la muestra preventiva y focalizada de sostenedores y/o escuelas para el programa uso de recursos SEP a ser enviada al proceso de gestión de las fiscalizaciones para efectuar el agendamiento y ejecución del servicio de fiscalización, oportunamente. Además, esta arquitectura, en la actividad planificar programas de fiscalización preventivos prepara acciones de prevención y/o mitigación (no solo visitas a terreno) formuladas para apoyar el logro de los objetivos del programa; las cuales, para el programa usos de recursos SEP, objeto del proyecto, estas se traducen en recomendaciones oportunas relacionadas con el uso de los fondos declarados por los sostenedores en las rendiciones de cuentas y que buscan, que estos rindan sus gastos adecuadamente (respetando la legalidad del uso de los recursos) y con ello, tener la seguridad que los gastos rendidos no serán objetados y/o reintegrados sus montos.

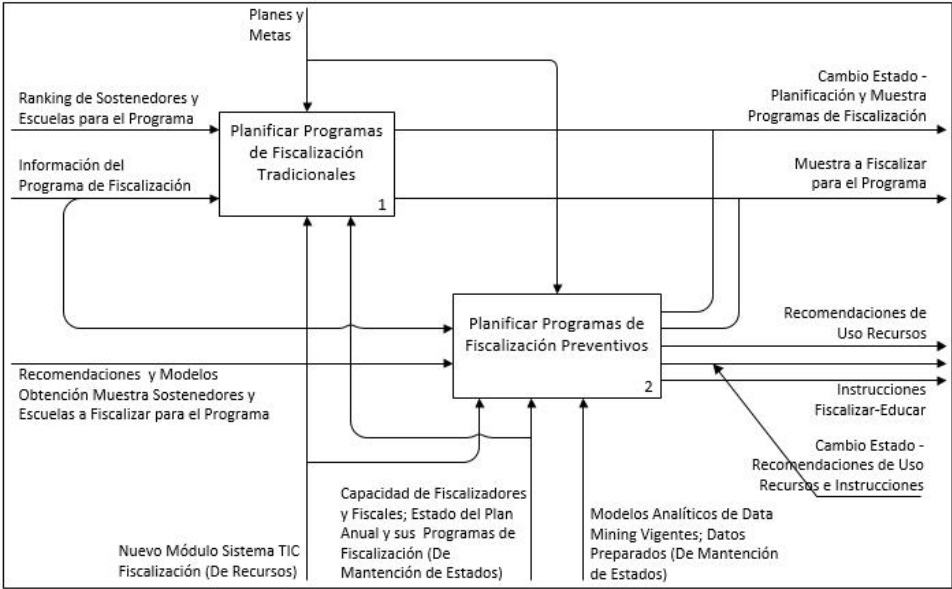


Figura N° 20: Planificar programas de fiscalización (to be)

Al instanciar el proceso planificar programas de fiscalización preventivos, se obtiene la arquitectura ilustrada en la Figura N° 21.

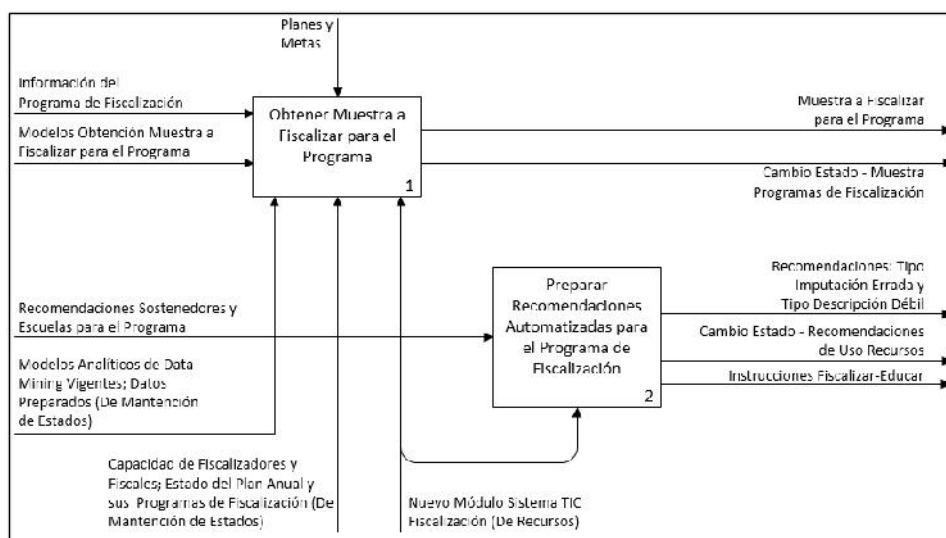


Figura N° 21: Planificar programas de fiscalización preventivos

El funcionamiento de esta arquitectura, para los propósitos del proyecto, instancia on demand la actividad obtener muestra a fiscalizar para el programa e instancia diariamente y durante todo el proceso de rendición de cuentas la actividad preparar recomendaciones automatizadas para el programa de fiscalización.

La actividad obtener muestra a fiscalizar para el programa, ejecuta los modelos analíticos vigentes para el programa uso de recursos SEP, preparados en el proceso desarrollar modelos de comportamientos de sostenedores y escuelas, y en función de sus resultados prepara la muestra a fiscalizar con aquellos sostenedores y/o escuelas que presenten la mayor probabilidad de estar incumpliendo con el debido uso de los fondos SEP. Los detalles de esta actividad se presentan en la Sección 5.2.2 Diseño en BPMN.

La actividad preparar recomendaciones automatizadas para el programa de fiscalización, ejecuta automática y diariamente el clasificador (diseñado para el proyecto) de textos a cada descripción de gasto declarado en el sistema de rendición de cuentas y cuyos resultados clasificados en: uso indebido, error de imputación en cuenta de gasto o escasa descripción del gasto; prepara recomendaciones automatizadas que serán incluidas en: el portal Web de rendición de cuentas o bien ser derivadas al proceso de gestión de las

fiscalizaciones para incorporarlas como un ítem en cada visita de fiscalización preventiva a terreno para fiscalizar-educar al sostenedor y/o escuela, o efectuar el seguimiento y control de subsanación al ser incluidas como una materia adicional en otros programas de fiscalización tradicionales (aprovechando la visita a terreno). Con esta forma de operar, se busca disminuir la cantidad de documentos riesgosos al final del proceso declarativo; evitando con ello, fiscalizaciones, impugnaciones administrativas de parte de sostenedores (defensa), esfuerzos de fiscales al decidir si son gastos no aceptados y esfuerzos de gestión interna para lograr el reintegro de la totalidad de los montos al sistema escolar en el próximo proceso declarativo. Los detalles de estas actividades se presentan en la Sección 5.2.2 Diseño en BPMN.

### **5.2.2 Diseño en BPMN**

Los siguientes diagramas detallan las actividades de los procesos requeridos para los propósitos del proyecto y donde, cada actividad representa la práctica de trabajo requerida para funcionar a nivel de roles, tareas, decisiones, reglas de negocios y flujos de datos. Cada práctica de trabajo, resalta las interacciones con el sistema tecnológico de apoyo propuesto por el proyecto y la lógica de negocio en caso de contenerla.

#### **5.2.2.1 Analizar Comportamientos de Sostenedores y Escuelas (To Be)**

Las siguientes actividades tienen por objeto diseñar y validar modelos analíticos, junto con descubrir patrones y recomendaciones para el programa uso de recursos SEP. Información que será utilizada en las actividades del proceso planificar programas de fiscalización preventivos.

##### **5.2.2.1.1 Preparar datos sostenedores y escuelas**

Tal como se ha descrito, para aplicar modelos analíticos se requiere construir una base de datos limpia; por ello, se debe trabajar en unificar diversas fuentes de información con distintos formatos y que pueden contener datos nulos, outliers, entre otros.

Esta actividad es instanciada con tres propósitos, el primero se hace cargo de la solicitud de preparar información requerida por los modelos analíticos de segmentación y clasificación de sostenedores y/o escuelas para preparar el programa de fiscalización uso de recursos SEP preventivo y focalizado (ver Figura N° 22); el segundo, se hace cargo

de preparar información para aplicar el modelo de clasificación de descripciones de gastos de las rendiciones de cuentas para preparar recomendaciones preventivas y/o mitigación para el mismo programa de fiscalización (ver Figura N° 23); el tercero, no cambia su funcionamiento respecto a su situación actual descrita en Sección 4.2.2.1 Preparar Ranking de Sostenedores y/o Escuelas.

La actividad ilustrada en la Figura N° 22, al ser instanciada encarga al analista de datos preparar solicitudes de información ad hoc a los propósitos del programa de fiscalización a ejecutar y las envía a entidades como: MINEDUC, Agencia de la Calidad de la Educación, SII entre otras. En paralelo, el analista de datos ejecuta (o prepara y ejecuta si es vez primera para el programa de fiscalización) las ETL's al repositorio de datos institucionales (DW). Con la información interna y externa recopilada, el analista de datos utiliza el sistema especializado (*hardware*: server dedicado y *software*: Stata y SPSS Modeler) para unificar los formatos de datos, tratar datos nulos y *outliers*; la información resultante se almacena en mantención de estados para su uso en posteriores procesos.

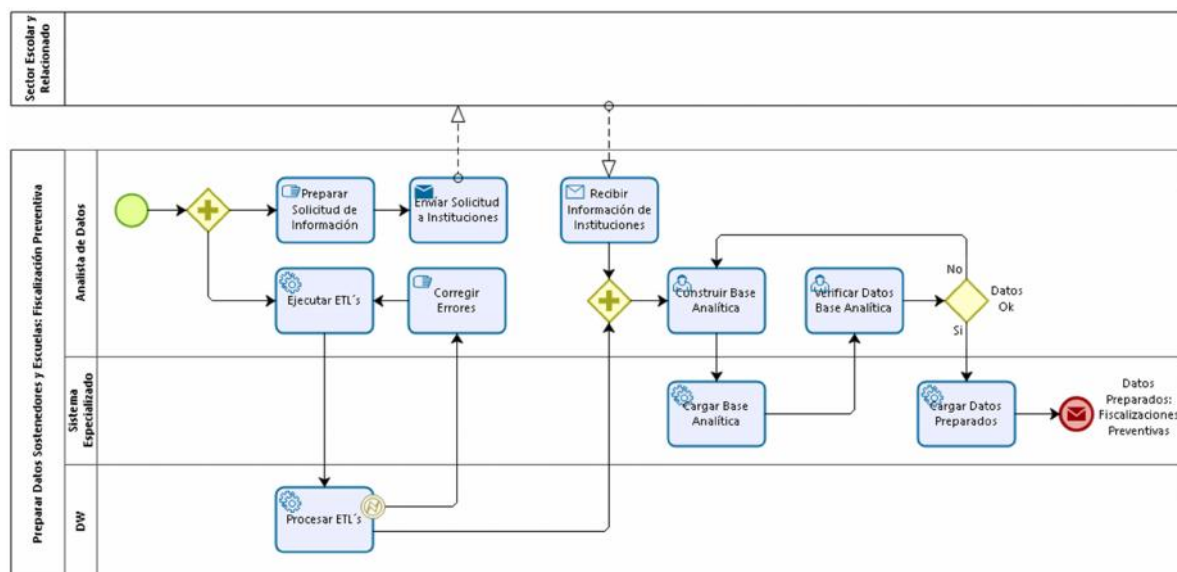


Figura N° 22: Preparar datos sostenedores y escuelas: fiscalización preventiva

La actividad ilustrada en la Figura N° 23, está programada por completo en el sistema TIC de apoyo propuesto por el proyecto, es instanciada diariamente a las 03:00 AM por un evento Jobs de SQL Server el cual activa una ETL a la base de datos del sistema de rendición de cuentas que ejecuta una SQL a sus tablas de datos y obtiene la información

para crear el archivo de texto (.csv) que será almacenado en mantención de estados para su uso en procesos posteriores. Esta actividad adicionalmente, cuenta con tareas automáticas que alertan al analista de datos si el proceso descrito presenta errores.

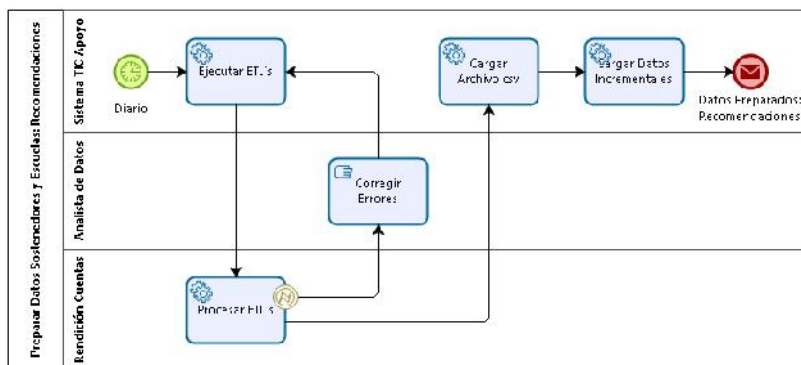


Figura N° 23: Preparar datos sostenedores y escuelas: recomendaciones

### 5.2.2.1.2 Desarrollar Modelos Comportamientos de Sostenedores y Escuelas

A continuación, se describen las actividades ilustradas en la Figura N° 18 a nivel de práctica de trabajo requerida para funcionar a nivel de roles, tareas, decisiones, reglas de negocios y flujos de datos. Cada práctica de trabajo, resalta las interacciones con el sistema tecnológico de apoyo propuesto por el proyecto y la lógica de negocio en caso de contenerla.

- Actividad: Exploración de datos

La actividad ilustrada en la Figura N° 24, es instanciada por el evento datos preparados proveniente de la actividad preparar datos sostenedores y escuelas: fiscalizaciones preventivas, el cual solicita al analista de negocio preparar un conjunto de criterios de consistencia y completitud a los datos preparados. Estos criterios se ejecutan con apoyo del sistema especializado y dependiendo de los resultados obtenidos, el analista decide si informar al analista de datos reprocesar los datos o almacenarlos en mantención de estados para el uso en siguientes procesos.



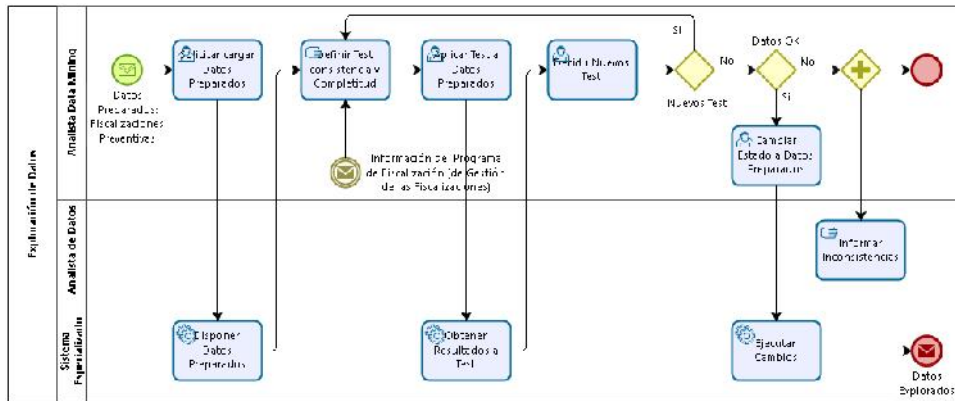


Figura N° 24: Exploración de datos

- Actividad: Selección y/o adecuación de variables a modelos

La actividad ilustrada en la Figura N° 25, es instanciada por el evento datos explorados proveniente de la actividad exploración de variables, el cual solicita al analista de negocio aplicar distintos test estadísticos a las variables y datos explorados. Estos test se ejecutan con apoyo del sistema especializado y dependiendo de los resultados obtenidos, el analista decide el conjunto de variables a utilizar y las almacena en mantención de estados para el uso en siguientes procesos.

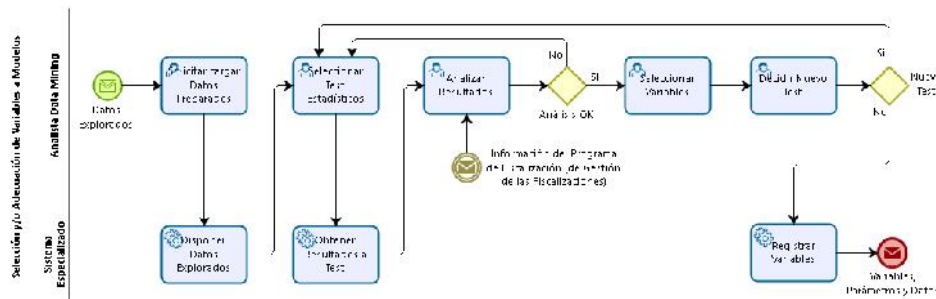


Figura N° 25: Selección y/o adecuación de variables a modelos

- Actividad: Elaboración de modelos de clustering sostenedores y/o escuelas

La actividad ilustrada en la Figura N° 26, es instanciada por el evento variables y datos, proveniente de la actividad selección y/o adecuación de variables a modelos, el cual solicita al analista de negocio en colaboración a los roles de fiscalización y responsable del programa elaborar modelos de clustering de sostenedores y/o escuelas para el programa de fiscalización en cuestión. Para lo anterior, el analista de negocio carga la

fuente de datos y variables almacenada en mantención de estados en el sistema especializado, en el cual selecciona el método de clustering a aplicar, un subconjunto de variables y previo a ejecutar el método, ajusta los parámetros requeridos por el método seleccionado. Los resultados obtenidos, numéricos y gráficos, son analizados por fiscalizadores expertos, analistas de negocio y el responsable del programa, con el propósito de encontrar patrones en los clústeres; este es un proceso iterativo (se repite tantas veces como sea necesario) se obtiene el número óptimo de clúster (en base a algún índice), se agregan o quitan variables, se ejecuta el método y se analiza las movilidades de los elementos en los clústeres (se busca que los clústeres tengan sentido de negocio). Los clústeres válidos, aquellos que explican los comportamientos de sostenedores y/o escuelas se clasifican con nombres representativos y se almacenan en mantención de estados para uso en procesos posteriores; en particular, a cada clúster de interés se le genera un modelo de clasificación para predecir su comportamiento y extraer las reglas que lo definen; estas reglas, en procesos sucesivos conformarán la(s) lógica(s) de negocio requeridas para los propósitos del proyecto.

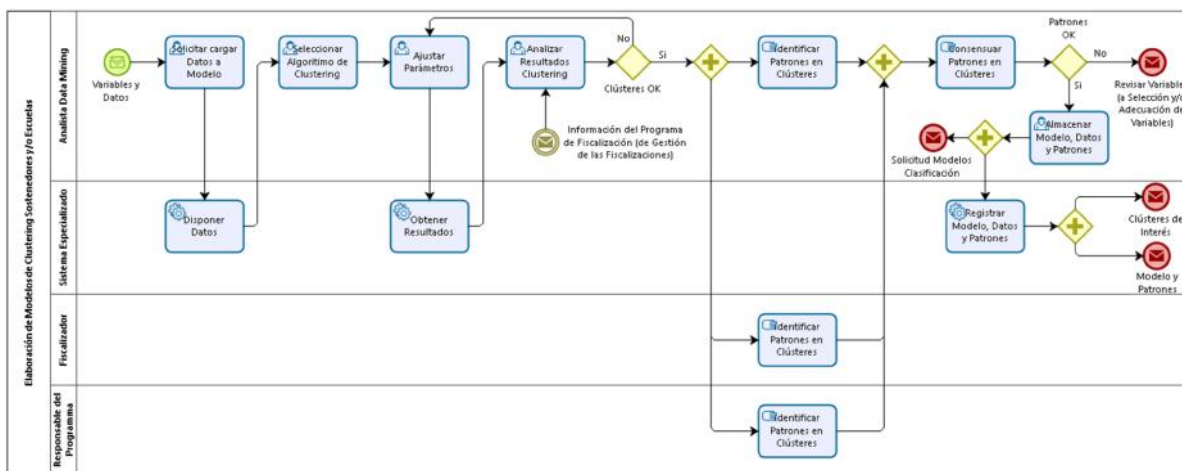


Figura N° 26: Elaboración de modelos de clustering sostenedores y/o escuelas

- Actividad: Elaboración de modelos de clasificación sostenedores y/o escuelas

La actividad ilustrada en la Figura N° 27, es instanciada por el evento solicitud de modelos proveniente de la actividad elaboración de modelos de clustering sostenedores y/o escuelas, el cual solicita al analista de negocio en colaboración a los roles de fiscalización y responsable del programa elaborar modelos de clasificación de sostenedores y/o

escuelas a los clústeres de interés para el programa de fiscalización en cuestión. A modo de resumen, los modelos de clasificación consisten en una serie de patrones útiles para distinguir clases, que tras ser entrenados (con una muestra de datos) pueden ser utilizados para predecir la clase de otros registros no clasificados.

Esta actividad una vez instanciada, el analista de negocio carga la fuente de datos (que incluye el atributo clúster de la actividad anterior) y las variables almacenada en mantención de estados en el sistema especializado, en el cual filtra el set de datos por el atributo clúster (seleccionado el clúster de interés), selecciona el subconjunto de variables (los mejores predictores para el atributo a predecir, resultantes de test estadísticos), divide el set de datos en función de una regla de partición en los conjuntos de entrenamiento y evaluación, selecciona el método de clasificación a aplicar, y previo a ejecutar el método, ajusta los parámetros requeridos por el método seleccionado. En primera instancia, los resultados obtenidos, numéricos y gráficos, son analizados por el analista de negocio, el cual decide si alterar el subconjunto de variables (incluye, quita, o construye) y/o modificar los parámetros del método seleccionado, este proceso es iterativo hasta lograr conformidad (calidad del modelo en función de algún indicador); a continuación, el analista de negocio repite el proceso con otro método de clasificación, en busca del mejor modelo para la problemática a predecir. Concluidos los ciclos descritos, los resultados, numéricos y gráficos, del modelo ganador son analizados por fiscalizadores expertos, analistas de negocio y el responsable del programa, con el propósito de encontrar patrones en forma de reglas, lógicas de negocio para los propósitos del proyecto, para una mejor comprensión por parte de las personas encargadas de interpretar los resultados. El modelo y patrones descubiertos se almacenan en mantención de estados para uso en procesos posteriores.

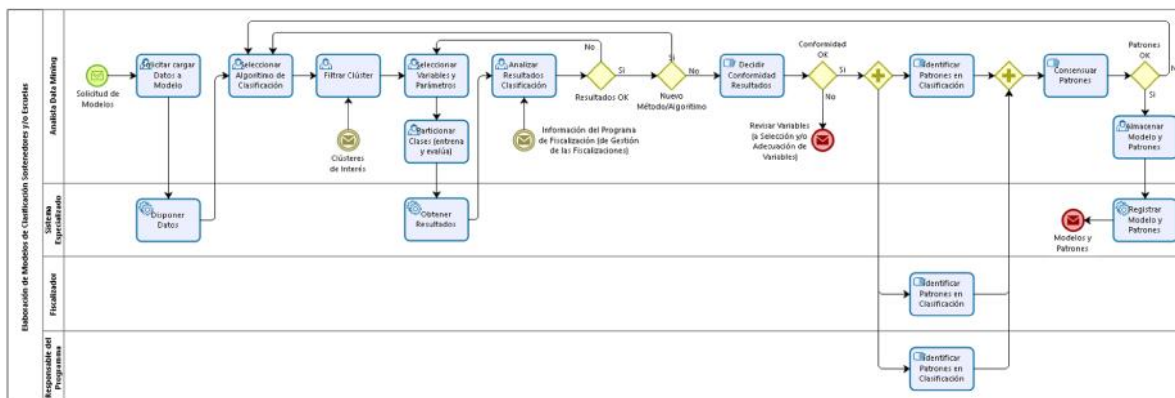


Figura N° 27: Elaboración de modelos de clasificación sostenedores y/o escuelas

- Actividad: Análisis de resultados y generación de patrones

Esta actividad ha sido descrita previamente en las actividades de elaboración de modelos de clustering y clasificación de sostenedores y/o escuelas.

### 5.2.2.1.3 Evaluar Modelos Comportamientos de Sostenedores y Escuelas

La actividad ilustrada en la Figura N° 28, es instanciada periódicamente por el evento solicitud de evaluación de modelos proveniente del proceso desarrollar modelos comportamientos de sostenedores y/o escuelas, el cual solicita al analista de negocio evaluar los modelos de clustering y clasificación de sostenedores y/o escuelas para el programa de fiscalización en cuestión. A modo de resumen, un modelo válido con el tiempo puede dejar de serlo, debidos a cambios en el sector (negocio) para el cual fue diseñado; por ejemplo, los sostenedores y/o escuelas pueden aprender las actuales formas de detección de incumplimientos y puede ser, que en vez de cambiar sus prácticas (para bien) decidan encontrar maneras diferentes de evadir o incumplir sin ser detectados.

Esta actividad una vez instanciada, el analista de negocio en flujos de tareas paralelas solicita a mantención de estados los modelos analíticos vigentes para el programa en cuestión y al sistema de fiscalización la información de resultados clasificados por el modelo, más nueva información del sector escolar y otra relacionada. Esta información preparada será la base para evaluar la efectividad de los modelos tras ser instanciados en el sistema especializado; dependiendo del margen de error (rango acordado) resultante de ejecutar los modelos, el analista de negocio decide mantener los modelos

o ajustarlos, en este último caso iterativamente se busca reducir el margen de error, en caso de lograrlo almacena los nuevos modelos en mantención de estados para uso en procesos posteriores o en caso contrario, informa la necesidad de nuevos modelos al proceso desarrollar modelos comportamientos de sostenedores y escuelas.

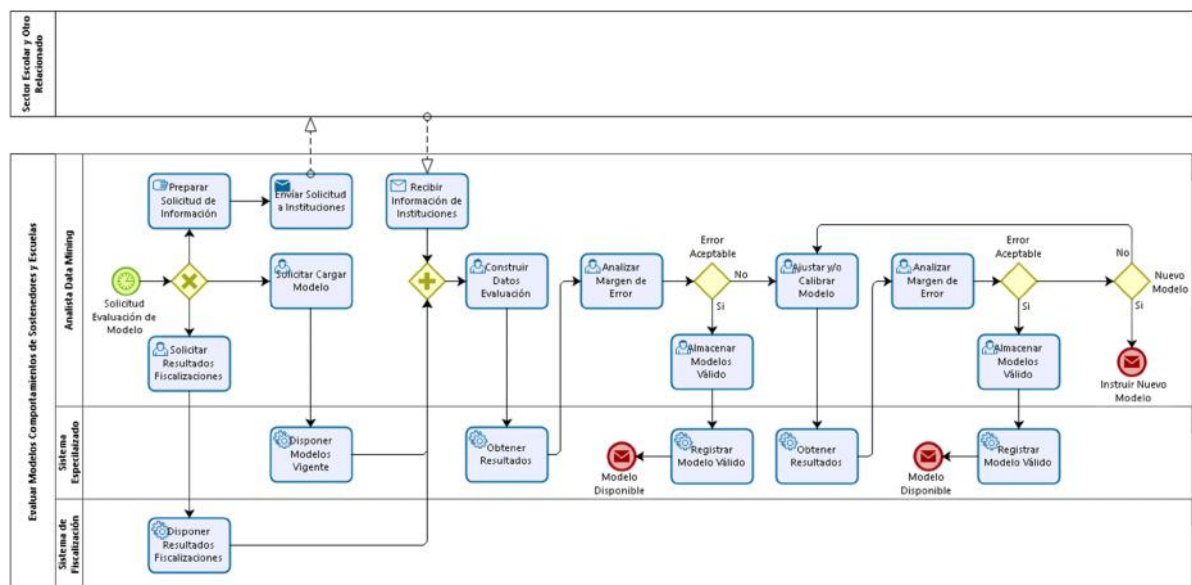


Figura N° 28: Evaluar modelos comportamientos de sostenedores y escuelas

#### 5.2.2.1.4 Desarrollar Modelos Obtención Muestra y Recomendaciones para el Programa

- Desarrollar Modelos Obtención Muestra para el Programa de Fiscalización

Tal como fue descrito esta actividad no cambia su funcionamiento respecto a su situación actual descrita en Sección 4.2.2.1 Preparar Ranking de Sostenedores y/o Escuelas.

- Desarrollar Recomendaciones para el Programa de Fiscalización

La actividad ilustrada en la Figura N° 29, está programada por completo en el sistema TIC de apoyo propuesto por el proyecto y es instanciada diariamente por el evento datos preparados: recomendaciones (archivo de texto .csv diarios con los detalles de gastos de las rendiciones de cuentas), el cual se carga automáticamente en el ambiente requerido por el componente que ejecutará el clasificador de textos a cada descripción de gasto (una fila de archivo de texto.csv). El resultado obtenido, es un archivo de texto con un atributo que indica por cada fila el tipo de recomendación (tipo

1: uso indebido; tipo 2: error de imputación en cuenta de gasto; tipo 3: escasa descripción del gasto), base para implementar la lógica de negocio requerida para los propósitos del proyecto. Las recomendaciones se almacenan en mantención de estados para uso en procesos posteriores. Esta actividad adicionalmente, cuenta con tareas automáticas que alertan al analista de datos o al analista desarrollador si el proceso descrito presenta errores.

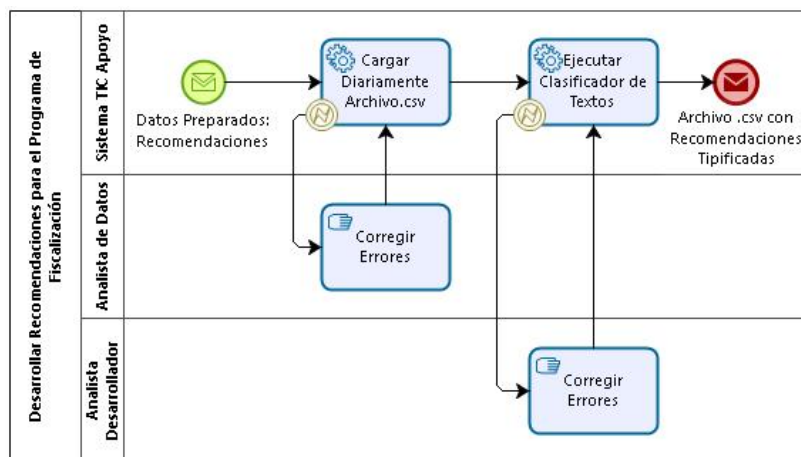


Figura N° 29: Desarrollar recomendaciones para el programa de fiscalización

### 5.2.2.2 Planificar Programas de Fiscalización (To Be)

Las siguientes actividades tienen por objeto ejecutar los modelos analíticos desarrollados para el programa uso de recursos SEP y dar utilidad a la información generada, ya sea para preparar fiscalizaciones preventivas focalizadas y/o recomendaciones individualizadas a sostenedores y/o escuelas.

#### 5.2.2.2.1 Planificar programas de fiscalización preventivos

- Obtener muestra a fiscalizar para el programa

La actividad ilustrada en la Figura N° 30, es instanciada por el evento modelos obtención muestra a fiscalizar para el programa, informando que los modelos se encuentran disponibles; conocida esta información, el jefe de fiscalización revisa los planes y metas para decidir la fecha en la cual aplicar el programa preventivo, y comunicarla al analista.

El analista de negocio debe determinar aquellos sostenedores y/o escuelas con la mayor probabilidad de estar incumpliendo con el debido uso de los fondos SEP y para ello,

recurre al sistema especializado para: cargar el set de datos, filtrar los clústeres de interés y ejecutar los modelos disponibles y válidos; los resultados obtenidos, se envían al jefe de fiscalización para decidir, apoyado de la lógica de negocio propuesta por el proyecto resaltada en la Figura N° 30, la muestra a fiscalizar para el programa.

Esta actividad, adicionalmente cuenta con otra lógica de negocio que apoya la toma de decisiones al jefe de fiscalización en aquellos casos donde el tamaño de la muestra a fiscalizar sea superior a la cantidad de visitas disponibles y donde, el problema no sea simple de resolver. En este caso, se debe priorizar el listado obtenido de sostenedores y/o escuelas, utilizando para ello parámetros como: cantidad de sostenedores y/o escuelas nunca fiscalizadas a incluir, cantidad mínima de sostenedores y/o escuelas municipales a incluir, una lista de sostenedores y/o escuelas fijas a incluir, entre otros; para ello, se recurre a un algoritmo de optimización matemática, desarrollado para los propósitos del proyecto, que es operado por el analista (iterativamente a solicitud del jefe de fiscalización). Con ello, se obtiene la muestra a fiscalizar que es almacenada en mantención de estados para ser utilizada en el proceso de gestión de las fiscalizaciones.

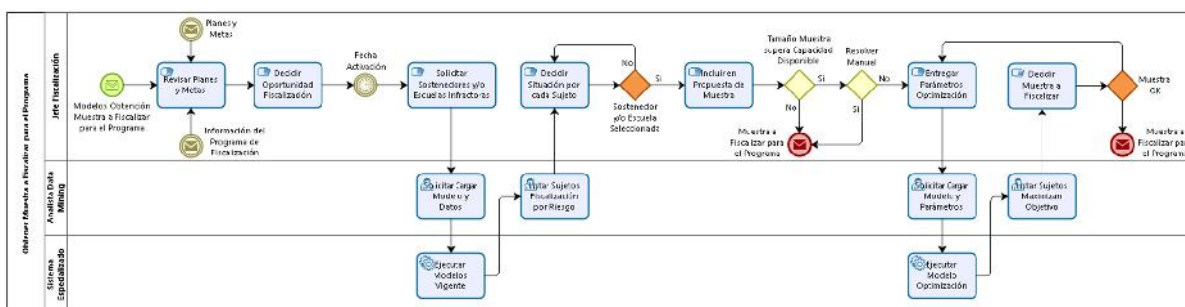


Figura N° 30: Obtener muestra a fiscalizar para el programa

- Preparar recomendaciones automatizadas para el programa de fiscalización

La actividad ilustrada en la Figura N° 31, está programada por completo en el sistema TIC de apoyo propuesto por el proyecto y es instanciada diariamente por el evento archivo .csv con recomendaciones tipificadas. Este archivo, es almacenado en un filesystem propio para ser procesado (leído) línea a línea. Por cada registro leído, se recurre a la lógica de negocio propuesta por el proyecto (resaltada en la Figura N° 31) que define las acciones a seguir en función del tipo de recomendación. Si la recomendación es de tipo 1 (uso indebido de recursos SEP), se almacena en

mantención de estados y se instruye al proceso de gestión de las fiscalizaciones incorporarla como una materia en cada visita de fiscalización preventiva a terreno con objeto de fiscalizar-educar al sostenedor y/o escuela o efectuar el seguimiento y control de subsanación de las mismas, al ser incluidas como una materia adicional en otros programas de fiscalización tradicionales (aprovechando la visita a terreno). Si la recomendación es de tipo 2 (error de imputación en cuenta de gasto) o de tipo 3 (escasa descripción del gasto), se consulta su pre existencia (si fue sugerida con anterioridad al sostenedor) en la base de datos del sistema de rendición de cuentas; de existir, se itera con una nueva fila del archivo de recomendaciones, de lo contrario, la recomendación se almacena en mantención de estados y es publicada en el sistema de rendición de cuentas (insertar nuevo registro a la base de datos). Recorridas todas las filas del archivo de texto de recomendaciones, se envía un mensaje a mantención de estados informando que el archivo fue procesado con éxito.

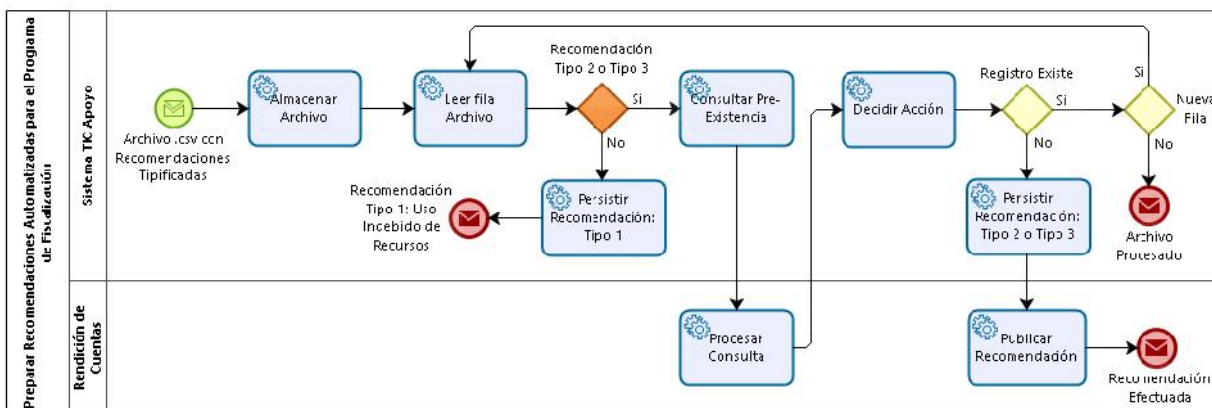


Figura N° 31: Preparar recomendaciones automatizadas para el programa de fiscalización



### 5.3 DISEÑO DE LÓGICA DE NEGOCIOS

Las lógicas de negocio identificadas para el proyecto pertenecen al proceso planificar programas de fiscalización preventivos y en particular, apoyan la toma de decisiones en las actividades obtener muestra a fiscalizar para el programa (ver Figura N° 30) y preparar recomendaciones automatizadas para el programa de fiscalización (ver Figura N° 31).

En la actividad obtener muestra a fiscalizar para el programa, coexisten dos lógicas negocio que comparten la responsabilidad de apoyar la toma de decisiones del jefe de fiscalización para definir la muestra preventiva a fiscalizar y focalizada en aquellos sostenedores y/o escuelas que tienen mayor probabilidad de incurrir en usos indebidos de recursos SEP. La primera lógica, para sugerir si un sostenedor y/o escuelas debe ser incluido en la muestra, recurre a la información descubierta de patrones y reglas provenientes de los resultados de modelos analíticos de segmentación y clasificación aplicados a incumplimientos financieros históricos de sostenedores y/o escuelas. La segunda lógica, opera sólo cuando el tamaño de la muestra a fiscalizar (resultante de la primera lógica) supera la capacidad de fiscalización definida para el programa y la problemática a resolver no es trivial; en este caso, para sugerir la muestra (subconjunto de elementos obtenido tras aplicar la primera lógica) se recurre la información resultante de emplear un algoritmo de programación matemática (optimización) que tras setear un conjunto de parámetros, define los elementos que maximizan el beneficio.

En la actividad preparar recomendaciones automatizadas para el programa de fiscalización, la lógica de negocio tiene la responsabilidad de sugerir la acción de prevención y/o mitigación adecuada a cada recomendación; para ello, recurre al modelo analítico de clasificación de textos, ideado para el proyecto, para asignar una categoría a cada descripción de gasto declarado por los sostenedores (durante todo el proceso de rendición de cuentas). Las categorías responden a: uso indebido, error de imputación en cuenta de gasto o escasa descripción del gasto; en función de ello, la lógica de negocio sugiere por cada recomendación, publicarla en el portal Web de rendición de cuentas o incorporarla como una materia más a fiscalizar en las visitas a terreno (independiente del programa) para asesorar-educar al sostenedor y/o escuela a subsanar situación.

### **5.3.1 Sugerir Muestra a Fiscalizar en Programa Preventivo Uso de Recursos SEP**

Tal como fue descrito, esta lógica sugiere si un sostenedor y/o escuelas debe ser incluido en la muestra a fiscalizar para el programa preventivo uso de recursos SEP; para ello, recurre a la información descubierta de patrones y reglas provenientes de los resultados de modelos analíticos de segmentación y clasificación aplicados a incumplimientos financieros históricos (entre los años 2014 a 2016) de sostenedores y/o escuelas.

- Entendimiento del negocio:

La SUPEREDUC anualmente debe fiscalizar el debido uso de fondos SEP a más de 7.900 escuelas (adscritas al convenio SEP) y en promedio a 2.270 de ellas, le detecta incumplimientos en sus rendiciones de cuentas e instruye a sus sostenedores el reintegro recursos. A juicio experto, no solo la rendición de cuentas refleja el uso indebido de estos fondos, y por ello, deben ser consideradas las siguientes situaciones: usar estos fondos para pagar las leyes sociales de sus trabajadores, en vez de utilizar la subvención general; saldar deudas producto de no pagar las leyes sociales de sus trabajadores al ser denunciados a los tribunales de justicia<sup>19</sup>; o preferir pagar la multa por no efectuar la declaración de gastos anual, en vez de realizar el proceso.

Detectar tempranamente los comportamientos riesgosos en esta materia y activar fiscalizaciones preventivas y focalizadas, es crucial para resguardar el debido uso de los recursos SEP -fondos que por Ley deben destinarse a apoyar a los estudiantes más vulnerables del país y a garantizar que los proyectos educativos se concreten-.

- Entendimiento de los datos:

Los datos a utilizar a priori, para los propósitos de la lógica de negocio, que caracterizan a las escuelas SEP son descritos en la dimensión (administrativa, geográfica, curriculares y gestión financiera) a la cual pertenecen.

Administrativos, el identificador único de la escuela o rol base de datos (RBD); el tipo de dependencia administrativa (municipal o particular subvencionado) del sostenedor

---

<sup>19</sup> Los tribunales ordenan al MINEDUC retener mensualmente un 20% del monto de la subvención general a aquellos sostenedores infractores, y hasta que estos demuestren haber saldado su deuda.

heredado a la escuela; la cantidad de cursos con matrícula impartidos durante un año; la cantidad de matrícula de alumnos educados durante un año; la cantidad de docentes contratados para prestar el servicio educativo durante un año; la cantidad de asistentes de la educación (no realizan clases en aula) contratados durante un año; la cantidad total de ingresos percibidos durante un año; la cantidad total de ingresos percibidos por la subvención general (con la cual se pagan las leyes sociales) durante un año; la cantidad total de gastos en sueldos de docentes y asistentes de la educación durante un año.

Geográficos, la región en la cual se ubica el local escolar y presta sus servicios; el tipo de localidad (urbana o rural) en la cual se ubica el local escolar.

Curriculares, el tipo de enseñanza (parvularia, básica, media humanista científica o media técnica profesional) que imparte la escuela; la asistencia promedio anual de los estudiantes matriculados; la tasa anual de aprobación o reprobación de estudiantes; la clasificación de la escuela (autónomo, emergente, recuperación) según resultados SIMCE y otras variables utilizadas por la Agencia de la Calidad de la Educación.

Gestión financiera, el monto de recursos SEP a reintegrar por la escuela tras ser detectados en la declaración anual de gastos de un año determinado; el monto total multado por la SUPEREDUC a la escuela en un año determinado; el monto total retenido por el MINEDUC (producto de no pago de leyes sociales) a la escuela en un año determinado; el tipo de apoyo de gestión financiera (servicio externalizado, personal propio o sin información) declarado por la escuela en un año determinado; un indicador que señala si la escuela rinde cuentas en un año determinado.

- Preparación de los datos:

Esta actividad es descrita en la Sección 5.2.2.1.1 Preparar datos sostenedores y escuelas y en función de ello, el analista de datos solicita a MINEDUC por cada escuela: el rut y dependencia administrativa del sostenedor, rol base de datos (RBD) de la escuela, los tipos de enseñanzas autorizados, la matrícula de alumnos informada, los cursos con matrícula impartidos, la asistencia promedio anual de los estudiantes matriculados, la tasa anual de aprobación o reprobación de estudiantes, los docentes y asistentes de la educación contratados, los recursos transferidos por subvención, la ubicación y zona del local escolar y el monto total retenido por no pago de leyes sociales; a la Agencia de la

Calidad de la Educación solicita por cada escuela: la clasificación asignada; mientras que, a los procesos internos de la SUPEREDUC solicita por cada escuela: el estado final de las rendiciones de cuentas, el monto de recursos SEP a reintegrar, el monto multado, el tipo de apoyo de gestión financiera declarado y el gasto en sueldos de docentes y asistentes de la educación informados en las rendiciones de cuentas.

Esta información, es almacenada en el *datawarehouse* de la SUPEREDUC (ver Figura N°32), implementado en una versión inicial como parte del alcance del proyecto; el cual facilita las labores del analista de datos de: extracción, transformación y carga de datos.

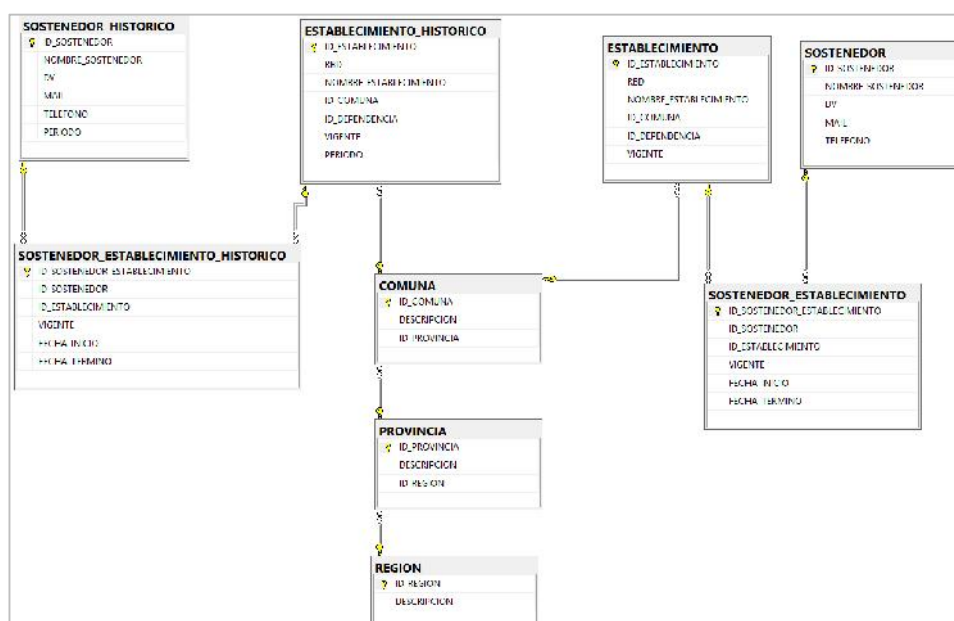


Figura N° 32: Extracto *datawarehouse* SUPEREDUC

A continuación, el analista de datos construye la base analítica limpia a la cual se aplicarán los modelos<sup>20</sup> de segmentación y clasificación; por ello debe, detectar y tratar: datos nulos, variables irrelevantes (alta concentración en algún valor), variables altamente correlacionadas y valores fuera de rango.

El listado de variables a ser sometidas a las tareas descritas, se presentan en la siguiente tabla de datos:

---

<sup>20</sup> El desarrollo de los modelos utiliza RapidMiner en su versión 5.3.015 y SPSS Modeler.

Tabla N° 3: Variables y valores

Variable	Valores
RBD	Entero > 0
Dependencia	Municipal o Particular Subvencionado
Región	Entero > 0
Zona	Urbano o Rural
Parvularia	0 o 1
Básica	0 o 1
Media HC	0 o 1
Media TP	0 o 1
Cursos	Entero > 0
Matrícula_2016	Entero > 0
Matrícula_2015	Entero > 0
Docentes_2016	Entero > 0
Docentes_2015	Entero > 0
Asistentes_2016	Entero > 0
Asistentes_2015	Entero > 0
Tasa_Asiencia_2016	Numérico > 0

Variable	Valores
Tasa_Aprobación_2016	Numérico > 0
Clasificación SEP	Autónomo, Emergente o En Recuperación
Ingresos_2016	Entero > 0
Gasto_Planta_2016	Entero > 0
Rinde_Cuentas_2016	0 o 1
Rinde_Cuentas_2015	0 o 1
Tipo_Apoyo	Servicio Externalizado, Personal Propio o Sin Dato
Gastos_Rechazados_2015	Entero > = 0
Gastos_Rechazados_2014	Entero > = 0
Multas_2014	Entero > = 0
Multas_2015	Entero > = 0
Multas_2016	Entero > = 0
Retenciones_2016	Entero > = 0
Retenciones_2015	Entero > = 0
Retenciones_2014	Entero > = 0

Datos nulos, son tratados utilizando el criterio *missing at random* (Little & Rubin,1987) debido a la relación existente entre datos nulos y los valores de otras variables; por ejemplo, casos donde la variable matrícula posee un valor distinto de 0 y la variable curso un valor 0, se tratan con la siguiente regla: redondear entero superior (matrícula/45). El resultado (columna *missings*) de este trabajo se presenta en la siguiente tabla.

Tabla N° 4: Resultado tratamiento datos nulos

Name	Type	Statistics	Range	Missings
RBD	integer	avg = 9170.943 +/- 7956.573	[1.000 ; 40459.000]	0
Dependencia	binominal	mode = Municipal (4889), least = Particular	Municipal (4889), Particular subv	0
Zona	binominal	mode = Urbano (4507), least = Rural (3432)	Urbano (4507), Rural (3432)	0
Clasificación	polynominal	mode = Emergente (6683), least = En Recu	Emergente (6683), Autónomo (1	0
Parvularia	binominal	mode = 1.0 (4859), least = 0.0 (3080)	0.0 (3080), 1.0 (4859)	0
Básica	binominal	mode = 1.0 (7156), least = 0.0 (783)	0.0 (783), 1.0 (7156)	0
Media HC	binominal	mode = 0.0 (6364), least = 1.0 (1575)	0.0 (6364), 1.0 (1575)	0
Media TP	binominal	mode = 0.0 (7544), least = 1.0 (395)	1.0 (395), 0.0 (7544)	0
Matrícula	integer	avg = 300.471 +/- 370.064	[1.000 ; 4129.000]	0
Cursos	integer	avg = 11.094 +/- 10.046	[1.000 ; 100.000]	0
Matrícula_Curso	numeric	avg = 22.650 +/- 12.259	[1.000 ; 102.000]	0
Matrícula_Ganancia	polynominal	mode = Mantiene (3378), least = Aumenta	Disminuye (2488), Mantiene (337	0
Docentes	integer	avg = 22.076 +/- 19.915	[1.000 ; 209.000]	0
Asistentes	integer	avg = 14.988 +/- 14.334	[0.000 ; 159.000]	0
Planta	integer	avg = 37.064 +/- 33.366	[1.000 ; 294.000]	0
Planta_Ganancia	polynominal	mode = Aumenta (3520), least = Disminuy	Aumenta (3520), Mantiene (3204	0
Gestión Financiera	polynominal	mode = Personal Propio (5321), least = Sir	Personal Propio (5321), Servic	0
Gasto Planta	numeric	avg = 356053410.756 +/- 363206941.130	[1006337.000 ; 3433629374.000]	0

Variables irrelevantes, aquellas con alta concentración en algún valor son identificadas utilizando herramientas de gráficos para posteriormente ser descartadas. Los detalles de cada variable de datos se encuentran disponibles en Anexo N° 3, por ejemplo, la Figura N° 33 ilustra la variable clasificación SEP y su alta concentración en el valor emergente.

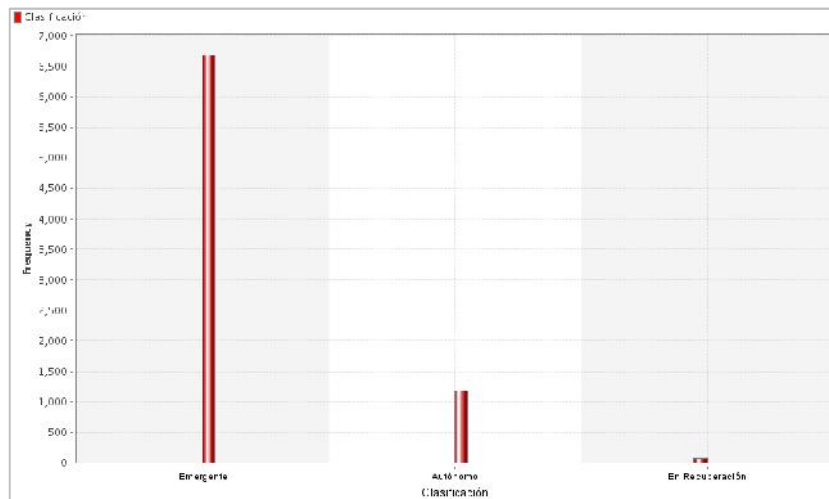


Figura N° 33: Concentración de valores en variable: Clasificación SEP

Las variables removidas por alta concentración en algún valor corresponden a: parvularia, básica, media humanista científica, media técnico profesional, tasa\_asistencia, tasa\_aprobación, clasificación\_SEP, rinde\_cuentas\_2016 y rinde\_cuentas\_2015.

Variables correlacionadas, se remueven todas aquellas variables que presenten una correlación sobre  $\pm 0,75$ . Para ello, se utiliza el proceso de la Figura N° 34; donde, el operador nominal a binominal transforma las variables nominales a numéricas (creando variables dummy) y el operador remover correlación quita aquellas variables correlacionadas respecto al valor definido en  $\pm 0,75$ . La tabla N° 5 ilustra el resultado de la matriz de correlación obtenida.

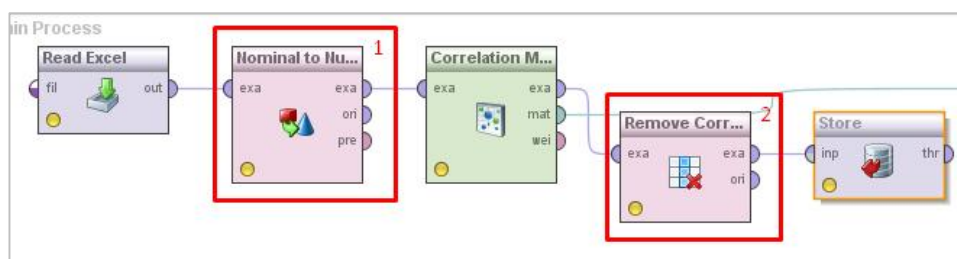


Figura N° 34: Proceso de identificación de variables correlacionadas

Las variables removidas por este proceso tienen relación a las variables dummy que en caso de tener dos valores debe dejarse solo una, tal es el caso de zona=urbana correlacionado con valor -1 con la variable zona=rural; y caso similar con las variables dummy para dependencia=municipal versus dependencia=particular subvencionado y

tipo\_apoyo =externalizado versus tipo\_apoyo=personal propio y tipo\_apoyo=sin dato; la otra variable con alta correlación corresponde a matrícula versus ingresos.

Tabla N° 5: Extracto de matriz de correlación

Atributos	Dependenc	Dependenc	Zona = Urb	Zona = Rural	Clasificació	Clasificació	Clasificació	Parvularia =	Parvularia =	Hásica = 0.0	Hásica = 1.0
Dependencia = Municipal	1	1	0.312	0.312	0.173	0.175	0.011	0.132	0.132	0.139	0.139
Dependencia = Particular subvencionado	-1	1	0.312	-0.312	-0.173	0.175	0.011	-0.132	0.132	0.139	-0.139
Zona = Urbano	-0.312	0.312	1	-1	-0.300	0.288	0.074	-0.416	0.416	0.251	-0.251
Zona = Rural	0.312	-0.312	-1	1	0.300	-0.288	-0.074	0.416	-0.416	-0.251	0.251
Clasificación = Emergente	0.173	-0.173	-0.300	0.300	1	-0.967	-0.218	0.286	-0.286	0.141	-0.141
Clasificación = Autónomo	-0.175	0.175	0.288	-0.288	-0.967	1	-0.040	-0.279	0.279	-0.137	0.137
Clasificación = En Recuperación	-0.011	0.011	0.074	-0.074	-0.218	-0.040	1	-0.050	0.050	-0.027	0.027
Parvularia = 0.0	0.132	-0.132	-0.416	0.416	0.286	-0.279	-0.050	1	-1	0.093	-0.093
Parvularia = 1.0	-0.132	0.132	0.416	-0.416	-0.286	0.279	0.050	-1	1	-0.093	0.093
Hásica = 0.0	-0.139	0.139	0.251	-0.251	0.141	-0.137	-0.027	0.093	-0.093	1	-1
Hásica = 1.0	0.139	-0.139	-0.251	0.251	-0.141	0.137	0.027	-0.093	0.093	-1	1
Media HC = 0.0	0.221	-0.221	-0.358	0.358	0.170	-0.180	0.020	0.031	-0.031	-0.110	0.110
Media HC = 1.0	-0.221	0.221	0.358	-0.358	-0.170	0.180	-0.020	-0.031	0.031	0.110	-0.110
Media TP = 1.0	0.010	0.010	0.134	-0.134	0.062	0.062	0.022	0.233	0.233	0.490	0.490
Media TP = 0.0	0.010	-0.010	-0.134	0.134	-0.062	-0.062	0.022	-0.233	0.233	-0.490	0.490

Valores fuera de rango, son identificados mediante herramientas de gráficos y la técnica de clustering con alto número de grupos. Un ejemplo del primer caso, corresponde a la variable gasto\_planta\_2016 presenta valor 0 y la suma de las variables docentes\_2016 + asistentes\_2016 es mayor que 0; donde, el valor 0 se reemplaza por el promedio<sup>21</sup>. Respecto a la técnica de clustering, se corre el modelo k-means para generar 25 grupos, y aquellos segmentos con la menor cantidad de casos (ver Figura N° 35) son removidos del conjunto de datos al ser considerados outliers multivariados.

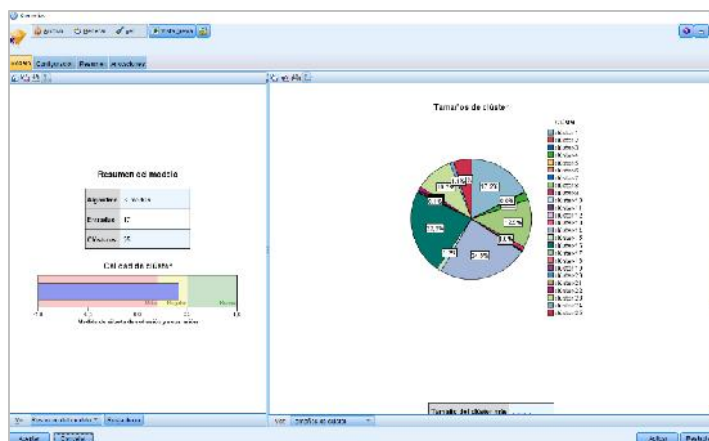


Figura N° 35: Detección de outliers multivariados

<sup>21</sup> Promedio corresponde a suma (total gasto planta2016) / suma (total docentes\_2016 + total asistentes\_2016).

Efectuado los procesos anteriores, adicionalmente se remueve la variable región por no aportar al objetivo buscado. Las variables resultantes se presentan en la siguiente tabla:

Tabla N° 6: Variables resultantes del proceso de limpieza

Variable	Valores	Variable	Valores
RBD	Entero > 0	Gasto_Planta_2016	Entero > 0
Dependencia_Municipal	0 o 1	Tipo_Apoyo_Externalizado	0 o 1
Región	Entero > 0	Gastos_Rechazados_2015	Entero > = 0
Cursos	Entero > 0	Gastos_Rechazados_2014	Entero > = 0
Matrícula_2016	Entero > 0	Multas_2014	Entero > = 0
Matrícula_2015	Entero > 0	Multas_2015	Entero > = 0
Docentes_2016	Entero > 0	Multas_2016	Entero > = 0
Docentes_2015	Entero > 0	Retenciones_2016	Entero > = 0
Asistentes_2016	Entero > 0	Retenciones_2015	Entero > = 0
Asistentes_2015	Entero > 0	Retenciones_2014	Entero > = 0
Ingresos_2016	Entero > 0		

Con los resultados obtenidos, se entiende que los datos están preparados, en conformidad a la actividad de Sección 5.2.2.1.1 Preparar datos sostenedores y escuelas.

- Modelamiento:

Este proceso es descrito en 5.2.2.1.2 Desarrollar Modelos Comportamientos de Sostenedores y Escuelas, y se rige por cada una de sus actividades.

La primera actividad de este proceso corresponde a explorar datos, el cual es realizado por el analista de data mining, y en gran parte de este trabajo fue presentada.

La segunda actividad de este proceso corresponde a seleccionar y/o adecuar variables y por lo cual, se crean las siguientes variables: variación\_matricula que señala si la escuela pierde o gana matrícula respecto al año anterior; matrícula\_curso indica la ocupación promedio de los cursos; planta que corresponde a la suma de docentes y asistentes; variación\_planta que señala si la escuela reduce o aumenta la planta respecto al año anterior; relación\_gasto\_ingreso señala la medida en que la planta se financia con los ingresos; multa como la suma de multa\_2016 y multa\_2015; retención como la suma de retenciones\_2016 y retenciones \_2015; y gastos\_rechazados como la suma de gastos\_rechazados\_2014 y gastos\_rechazados \_2015.

La siguiente tabla N° 7, resume la actividad seleccionar y/o adecuar variables a las cuales aplicar los modelos de segmentación y clasificación, en busca de patrones y reglas que expliquen comportamientos riesgosos en el debido uso de fondos SEP.



Tabla N° 7: Variables a modelos analíticos

Variable	Valores	Variable	Valores
RBD	Entero > 0	Variación_Planta	Numérico
Dependencia_Municipal	0 o 1	Ingresos_2016	Entero > 0
Zona_Urbano	0 o 1	Relación_Gasto_Ingreso	Numérico
Cursos	Entero > 0	Tipo_Apoyo_Personal_Propio	0 o 1
Matrícula_2016	Entero > 0	Gastos_Rechazados	Entero > = 0
Matrícula_Curso	Numérico	Multas	Entero > = 0
Variación_Matrícula	Numérico	Retenciones	Entero > = 0
Planta	Entero > 0		

La tercera actividad de este proceso corresponde a elaboración del modelo de clustering sostenedores y/o escuelas, que busca crear grupos donde los elementos pertenecientes a un grupo sean similares entre sí y diferentes respecto a los elementos de otros clústeres; para lo anterior, utiliza las variables de la Tabla N° 7; excluyendo las variables multas y retenciones, que se utilizarán para analizar la calidad de los clústeres descubiertos (utilizando el método criterio de negocio) y caracterizar a cada clúster.

La Figura N° 36 ilustra el modelo analítico utilizado y donde, el operador selección de atributos carga las variables a utilizar; el operador normalización establece un rango para todas aquellas variables numéricas; el operador clustering aplica el método k-means que asigna a cada escuela a un único clúster; y el operador performance que calcula el índice Davies Bouldin requerido para determinar el número óptimo de clúster.

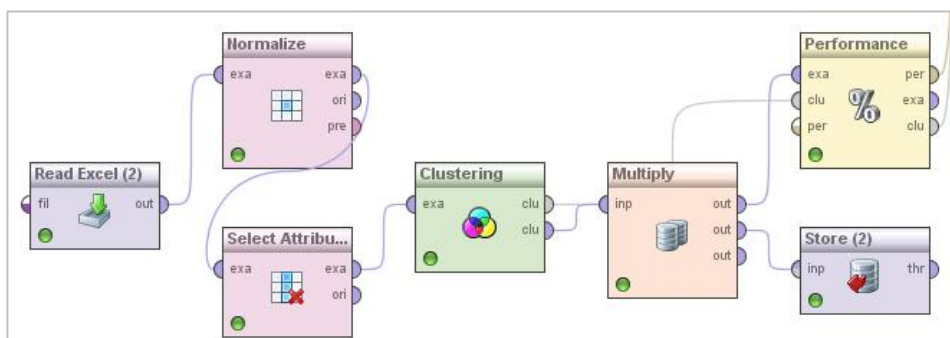


Figura N° 36: Proceso del modelo analítico de clustering

Tras iterar con distintos números de clústeres (k), el óptimo se obtiene con k= 5 con un índice Davies Bouldin = -1,07 y con la siguiente composición o cantidad de elementos: clúster\_0 = 855; clúster\_1 = 2.610; clúster\_2 = 2419; clúster\_3 = 93; clúster\_4 = 1962.

La siguiente figura ilustra los resultados del proceso iterativo para determinar el número óptimo de clústeres.

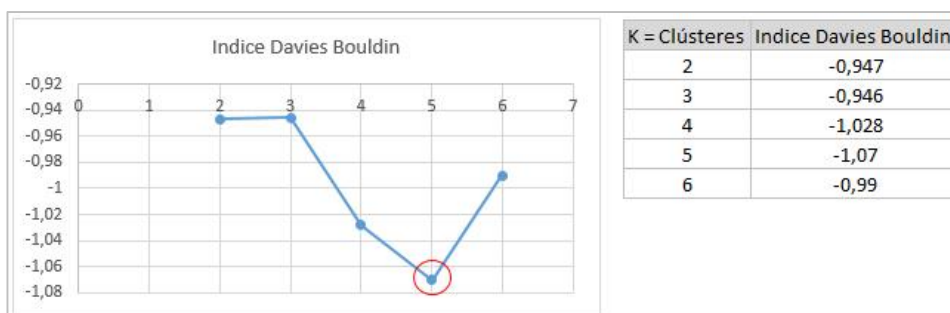


Figura N° 37: Índices Davies Bouldin para distintos números de clústeres

La siguiente tabla de datos ilustra los centroides de cada clúster.

Tabla N° 8: Tabla de centroides de clústeres

Attribute	cluster_0	cluster_1	cluster_2	cluster_3	cluster_4
Cursos	0.279	0.056	0.031	0.540	0.153
Matrícula_2016	0.235	0.039	0.008	0.499	0.114
Matrícula_Curso	0.326	0.259	0.065	0.364	0.283
Variación_Matrícula	0.051	0.053	0.051	0.051	0.051
Planta	0.316	0.082	0.026	0.563	0.192
Variación_Planta	0.147	0.151	0.161	0.145	0.148
Relación_Gasto_Ingreso	0.065	0.072	0.071	0.061	0.072
Gastos_Rechazados	0.047	0.012	0.003	0.064	0.026

Para complementar la decisión del número óptimo y calidad de los clústeres se recurre al criterio de negocio, que consiste en analizar la movilidad de los elementos de cada clúster al incluir una nueva variable de negocio; es decir, si una variable de negocio tiene relación con el problema a modelar, los clústeres deben en gran medida mantenerse estables.

En este caso, se cuenta con dos variables de incumplimientos, las retenciones que tienen relación directa con la gestión financiera y las multas, que en su mayoría responden incumplimientos normativos y no tanto con la gestión financiera.

La siguiente tabla de datos ilustra la movilidad de los elementos entre los clústeres al incluir la variable retenciones.

Tabla N° 9: Gastos\_rechazados versus retenciones

Gastos_Rechazados	Retenciones					Total
	cluster_0	cluster_1	cluster_2	cluster_3	cluster_4	
cluster_0	830			25		<b>855</b>
cluster_1		2.508			102	<b>2.610</b>
cluster_2		16	2.403			<b>2.419</b>
cluster_3				93		<b>93</b>
cluster_4	54				1.908	<b>1.962</b>
<b>Total</b>	<b>884</b>	<b>2.524</b>	<b>2.403</b>	<b>118</b>	<b>2.010</b>	<b>7.939</b>

De la tabla anterior, se aprecia que los elementos de los clústeres originales, al incluir la variable retenciones, no sufren mayores movimientos; es decir, los clústeres son consistentes con esta variable de negocio.

La siguiente tabla de datos ilustra la movilidad de los elementos entre los clústeres al incluir la variable multas.

Tabla N° 10: Gastos rechazados versus multas

Gastos_Rechazados	Multas					Total
	cluster_0	cluster_1	cluster_2	cluster_3	cluster_4	
cluster_0			127	728		<b>855</b>
cluster_1	2.585		10		15	<b>2.610</b>
cluster_2	3				2.416	<b>2.419</b>
cluster_3		48		45		<b>93</b>
cluster_4	157		1.804	1		<b>1.962</b>
<b>Total</b>	<b>2.745</b>	<b>48</b>	<b>1.941</b>	<b>774</b>	<b>2.431</b>	<b>7.939</b>

De la tabla anterior, se aprecia que los elementos de los clústeres originales, al incluir la variable multas, sufren movimientos; es decir, los clústeres no son consistentes con esta variable de negocio.

En conclusión, los clústeres obtenidos tienen sentido de negocio, el modelo es válido y sus variables son descritas en la Tabla N° 7, excluyendo sólo la variable multas.

A continuación, comienza el proceso de caracterización de cada clúster y que permita identificar aquellos clústeres riesgosos, para los propósitos del proyecto.

La siguiente tabla de datos caracteriza los clústeres en función de las variables: Dependencia\_Municipal, Zona\_Urbano, Tipo\_Apoyo\_Personal\_Propio y Matrícula\_2016.

Tabla N° 11: Caracterización versus primer set de variables

Clúster	Tamaño (clúster)	Cantidad			Suma
		Dependencia_Municipal	Zona_Urbano	Tipo_Apoyo_Personal_Propio	Matrícula_2016
cluster_0	855	398	838	573	828.671
cluster_1	2.610	1.501	1.370	1.622	419.501
cluster_2	2.419	1.777	380	1.778	83.936
cluster_3	93	25	89	59	191.591
cluster_4	1.962	1.188	1.830	1.289	925.253
<b>Total</b>	<b>7939</b>	<b>4.889</b>	<b>4.507</b>	<b>5.321</b>	<b>2.448.952</b>

De la tabla anterior, por ejemplo, se deduce que el clúster\_4 es aquel que cuenta con la mayor cantidad de estudiantes, y estos asisten en su mayoría a escuelas municipales.

La siguiente tabla de datos caracteriza los clústeres en función de las variables: Gastos\_Rechazados y Retenciones.

Tabla N° 12: Caracterización versus segundo set de variables

Clúster	Tamaño (clúster)	Cantidad		Total
		Gastos_Rechazados	Retenciones	
cluster_0	855	516	279	795
cluster_1	2.610	1.296	553	1.849
cluster_2	2.419	798	419	1.217
cluster_3	93	54	32	86
cluster_4	1.962	1.202	592	1.794
<b>Total</b>	<b>7939</b>	<b>3.866</b>	<b>1.875</b>	<b>5.741</b>

De la tabla anterior, por ejemplo, se deduce que el clúster\_1 es aquel que cuenta con la mayor cantidad de incumplimientos.

La siguiente tabla de datos caracteriza los clústeres en función de las variables numéricas en función de su valor promedio y desviación estándar.

Tabla N° 13: Caracterización versus tercer set de variables

Clúster	Tamaño (clúster)	Planta		Cursos		Matrícula		Matrícula_Curso	
		Promedio	Desviación ST	Promedio	Desviación ST	Promedio	Desviación ST	Promedio	Desviación ST
cluster_0	855	93,6	8,8	28,6	5,3	969,2	226,7	33,9	5,0
cluster_1	2.610	25,0	18,1	6,5	3,9	160,7	94,9	27,2	8,9
cluster_2	2.419	8,7	42,0	4,1	3,3	34,7	43,2	7,6	5,3
cluster_3	93	165,9	7,0	54,5	15,0	2.060,1	630,2	37,7	3,9
cluster_4	1.962	57,4	10,4	16,2	3,9	471,6	132,2	29,5	6,0
<b>Total</b>	<b>7939</b>								

De la tabla anterior, por ejemplo, se deduce que el clúster\_3 es aquel donde en promedio sus escuelas cuentan con el mayor número de cursos impartidos, y estos a su vez tienen el mayor indicador de aprovechamiento con un 37,7 (matrícula por curso).

La siguiente tabla de datos se utilizará para caracterizar los clústeres por grado de riesgo.

Tabla N° 14: Matrícula potencial afectada por incumplimientos

Clúster	Tamaño Clúster (1)	Cantidad Incumplimientos (2)	Tasa Incumplimiento (3) = (2)/(1)	Matrícula Total (4)	Matrícula Potencial Afectada (3)*(4)
cluster_0	855	795	0,93	828.671	770.519
cluster_1	2.610	1.849	0,71	419.501	297.187
cluster_2	2.419	1.217	0,50	83.936	42.228
cluster_3	93	86	0,92	191.591	177.170
cluster_4	1.962	1.794	0,91	925.253	846.026

Para los propósitos del proyecto, se caracterizarán los clústeres de mayor a menor nivel de riesgo en función de la matrícula potencial afectada (ver Tabla N° 14) y su descripción en función de las variables descritas en las Tablas N° 9, 10, 11, 12 y 13.

Clúster con riesgo muy alto, denominado los 14 cursos por generación, corresponde al cluster\_4 compuesto por 1.962 escuelas, el cual cuenta con 1.794 incumplimientos entre gastos\_rechazados y retenciones (la segunda mayor entre los clústeres); sus escuelas en total atienden 925.253 estudiantes (en promedio 470 estudiantes), en un 70% son municipales y su gestión financiera es realizada en un 60% con personal propio (escuela promedio o típica de parvularia a media).

Clúster con riesgo alto, denominado de dos generaciones, corresponde al cluster\_0 compuesto por 855 escuelas, el cual cuenta con 795 incumplimientos entre gastos\_rechazados y retenciones (la cuarta mayor entre los clústeres); sus escuelas en total atienden 828.671 estudiantes (en promedio 970 estudiantes), en un 50% son municipales y su gestión financiera es realizada en un 70% con personal propio (escuela mediana a grande).

Clúster con riesgo medio, denominado de un nivel a media capacidad, corresponde al cluster\_1 compuesto por 2.610 escuelas, el cual cuenta con 1.849 incumplimientos entre gastos\_rechazados y retenciones (la segunda mayor entre los clústeres); sus escuelas en total atienden 419.501 estudiantes (en promedio 160 estudiantes), en un 60% son

municipales y su gestión financiera es realizada en un 60% con personal propio (escuela pequeña).

Clúster con riesgo medio a bajo, denominado los grandes optimizados, corresponde al cluster\_3 compuesto por 93 escuelas, el cual cuenta con 86 incumplimientos entre gastos\_rechazados y retenciones (la menor entre los clústeres); sus escuelas en total atienden 191.591 estudiantes (en promedio 2.060 estudiantes), en un 70% son particulares subvencionadas y su gestión financiera es realizada en un 60% con personal propio (escuela grande).

Clúster con riesgo bajo, denominado los pequeños, corresponde al cluster\_2 compuesto por 2.610 escuelas, el cual cuenta con 1.217 incumplimientos entre gastos\_rechazados y retenciones (la tercera mayor entre los clústeres); sus escuelas en total atienden 83.936 estudiantes (en promedio 35 estudiantes), en un 70% son municipales y su gestión financiera es realizada en un 70% con personal propio (escuela muy pequeña).

Descubierto los patrones en los clústeres, en función de su nivel de riesgo y caracterización, finaliza la actividad elaboración del modelo de clustering de sostenedores y/o escuelas; se persisten los datos y se solicita a la actividad siguiente descubrir conocimiento en aquellos clústeres de interés.

La cuarta actividad de este proceso corresponde a la elaboración del modelo de clasificación sostenedores y/o escuelas, el cual tiene por objeto explicar el comportamiento del clúster de interés por medio de patrones y reglas. En este caso, se selecciona el clúster con riesgo muy alto (cluster\_4) compuesto por 1.962 escuelas.

La siguiente Figura N° 38 ilustra el modelo preparado para fines descritos.

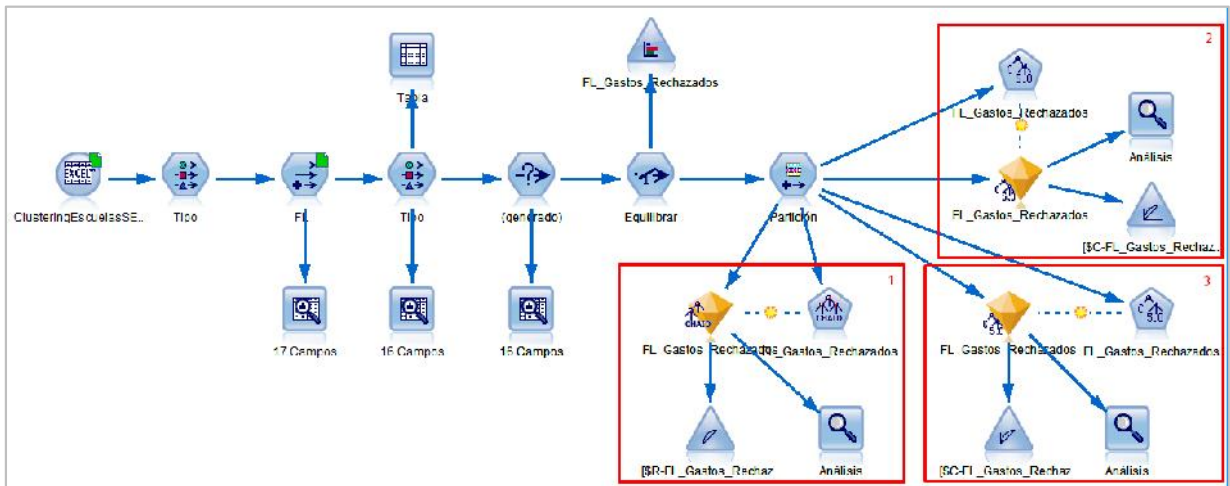


Figura N° 38: Proceso del modelo analítico de clasificación

La primera tarea del modelo, carga el set de datos resultante del proceso de clustering (operador leer Excel, clustering\_escuelas\_SEP) y tal como ilustra la Figura N° 39, incorpora la variable k\_óptimo.

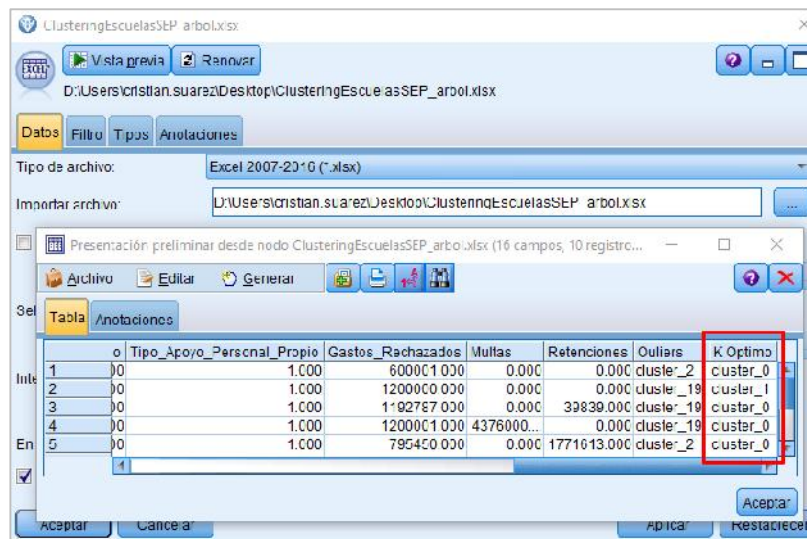


Figura N° 39: Carga de datos a modelo de clasificación

La siguiente tarea de esta actividad corresponde a la selección de los métodos o algoritmos de clasificación. Se escogen métodos de árboles de decisión, debido a que los resultados pueden ser visualizados gráficamente por roles no especializados en esta materia (fiscalizadores y el responsable del programa); por lo anterior, la problemática utilizará un árbol CHAID (Figura N° 38, punto 1) y un árbol C5.0 (Figura N° 38, punto 2).

La siguiente Figura N° 40 ilustra el proceso de filtrado de datos respecto al clúster de interés, clúster\_4 con 1.962 escuelas.

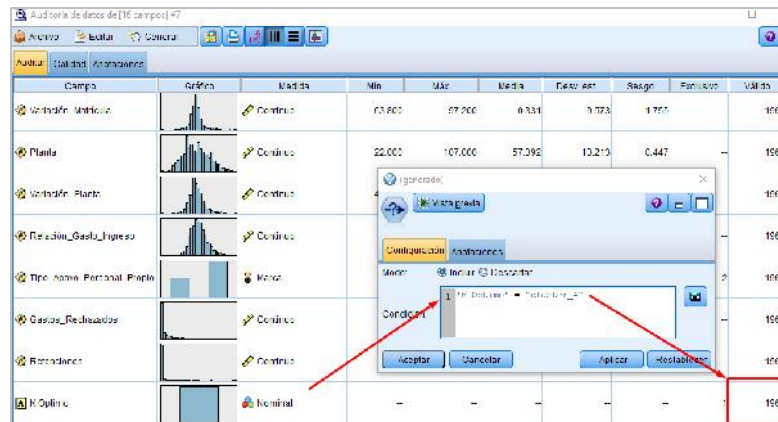


Figura N° 40: Filtrar set de datos por clúster de interés

A continuación, se selecciona la variable objetivo (label) que corresponde a gastos\_rechazados, previamente transformada a binominal (marca 0 o 1). Esta variable se explora para conocer la proporción de incumplimientos (marca en 1), tal como ilustra la Figura N° 41; debido a la alta proporción de valores en 1 (objetivo del modelo), se utiliza el operador equilibrar con la regla de 1 a 4 (tras iterar diversos valores) para aumentar la cantidad de valores 0 y obtener una proporción de 70% valores en 0 y 30% valores en 1.

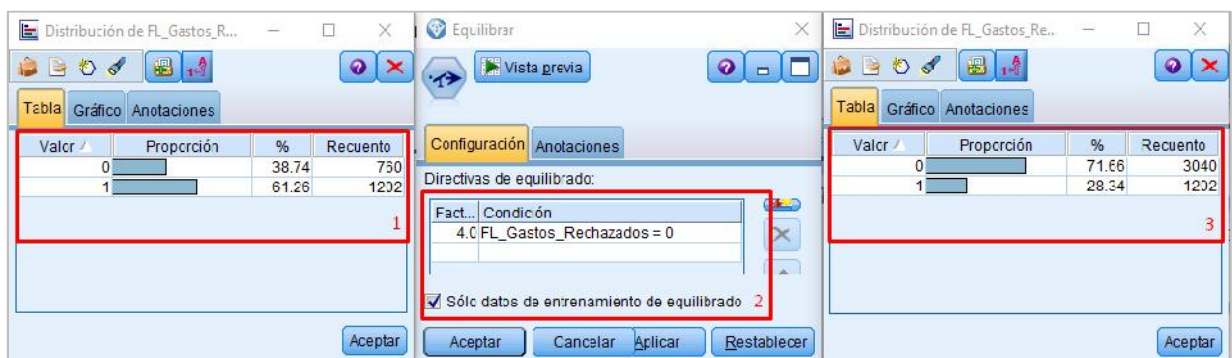


Figura N° 41: Proceso de equilibrado de la variable objetivo del modelo

Adicionalmente, se buscan los mejores predictores para la variable objetivo; para ello, se aplican test estadísticos; siendo los resultados obtenidos ilustrados en la Figura N° 42.



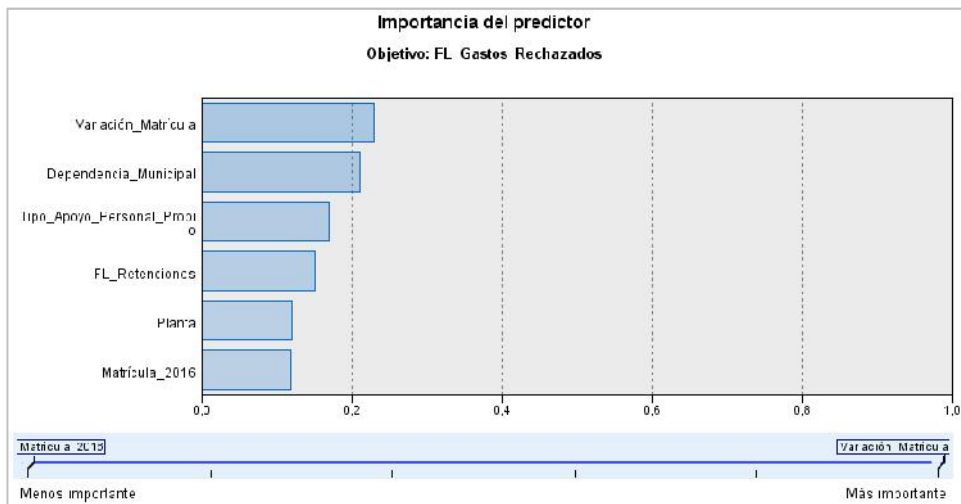


Figura N° 42: Importancia del predictor

Realizadas las tareas previas, se particionan las clases en los conjuntos de entrenamiento y test; para ello, se recurre al operador partición (ver Figura N°38) seteado en 80% entrenamiento y 20% test, tal como ilustra la Figura N° 43.

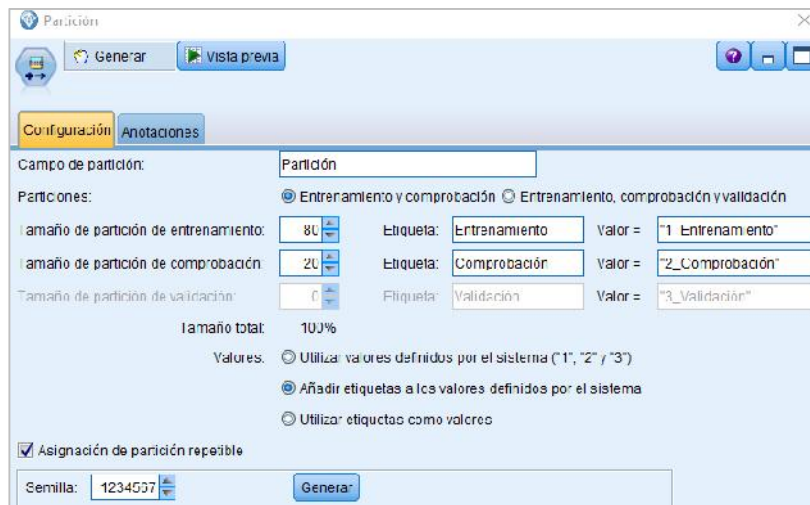


Figura N° 43: Proceso de partición del set de datos

La ejecución del modelo de clasificación de sostenedores y/o escuelas, ilustrado en la Figura N° 38, es un proceso iterativo al agregar/sacar variables y/o ajustar parámetros a los métodos o algoritmos de árboles CHAID y C5.0.

La siguiente Figura N° 44 ilustra el mejor resultado obtenido con árboles CHAID.

Resultados para el campo de resultado FL\_Gastos\_Rechazados

Modelos individuales

Comparando \$R-FL\_Gastos\_Rechazados con FL\_Gastos\_Rechazados

'Partición'	1_Entrenamiento		2_Comprobación	
Correctos	2.438	72,19%	606	70,06%
Erróneos	939	27,81%	259	29,94%
Total	3.377		865	

Matriz de coincidencias para \$R-FL\_Gastos\_Rechazados (las filas muestran las reales)

'Partición' = 1_Entrenamiento		0	1
0		2.400	25
1		914	38

'Partición' = 2_Comprobación		0	1
0		604	11
1		248	2

Evaluación del rendimiento

'Partición' = 1_Entrenamiento		
0		0,008
1		0,761

'Partición' = 2_Comprobación		
0		-0,007
1		-0,174

Métricas de evaluación

'Partición'	1_Entrenamiento		2_Comprobación	
Modelo	AUC	Gini	AUC	Gini
\$R-FL_Gastos_Rechazados	0,637	0,275	0,581	0,161

Figura N° 44: Mejor resultado obtenido con árboles CHAID

La siguiente Figura N° 45 ilustra el mejor resultado obtenido con árboles C5.0.

Resultados para el campo de resultado FL\_Gastos\_Rechazados

Modelos individuales

Comparando \$C-FL\_Gastos\_Rechazados con FL\_Gastos\_Rechazados

'Partición'	1_Entrenamiento		2_Comprobación	
Correctos	2.937	86,97%	751	86,82%
Erróneos	440	13,03%	114	13,18%
Total	3.377		865	

Matriz de coincidencias para \$C-FL\_Gastos\_Rechazados (las filas muestran las reales)

'Partición' = 1_Entrenamiento		0	1
0		2.416	9
1		431	521

'Partición' = 2_Comprobación		0	1
0		612	3
1		111	139

Evaluación del rendimiento

'Partición' = 1_Entrenamiento		
0		0,167
1		1,249

'Partición' = 2_Comprobación		
0		0,174
1		1,22

Métricas de evaluación

'Partición'	1_Entrenamiento		2_Comprobación	
Modelo	AUC	Gini	AUC	Gini
\$C-FL_Gastos_Rechazados	0,87	0,739	0,854	0,707

Figura N° 45: Mejor resultado obtenido con árboles C5.0

Para seleccionar el método con el cual analizar el comportamiento del clúster de interés, se recurre al indicador accuracy, seleccionando aquel que presente mayor valor. Dado lo anterior, el árbol C5.0 logra un accuracy de 0,85 versus un 0,58 para el árbol CHAID.

A continuación, se presenta la rama del árbol C5.0 seleccionada y consensuada para ser la regla requerida para los propósitos del proyecto, y ser utilizada en el proceso de la Sección 5.2.2.2.1 Planificar programas de fiscalización preventivos.

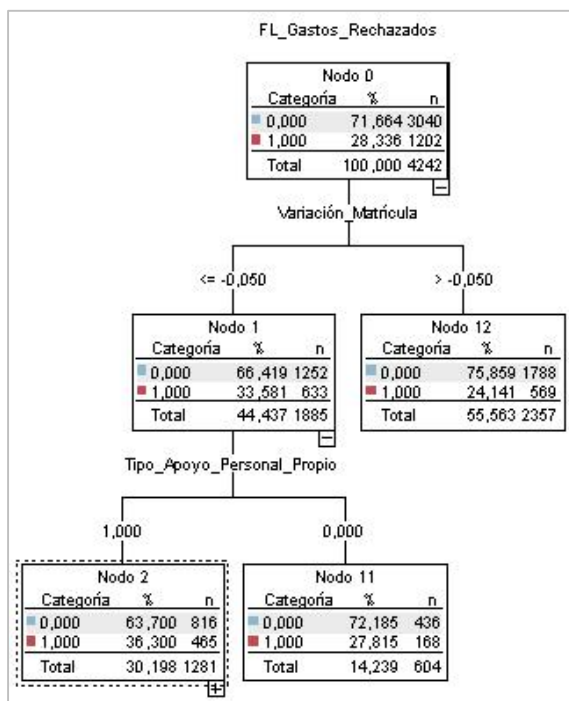


Figura N° 46: Rama árbol C5.0 seleccionada

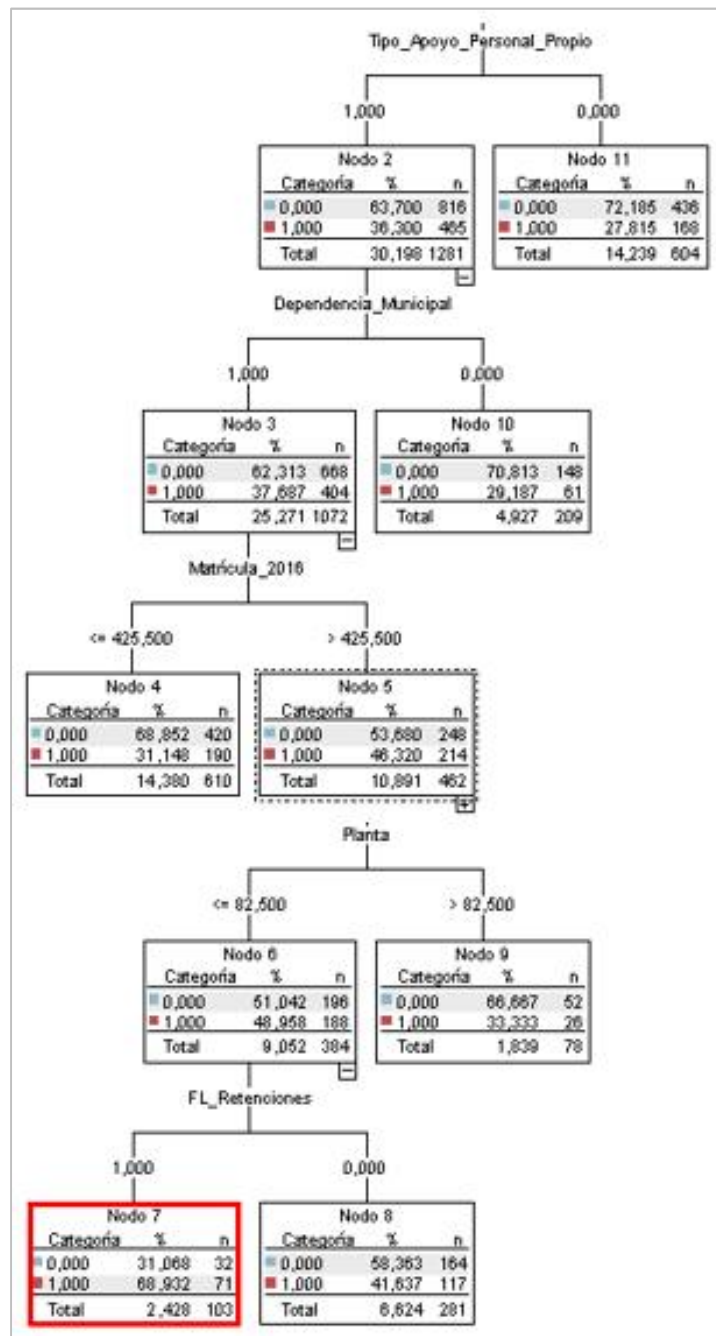


Figura N° 47: Rama árbol C5.0, continuación

De las Figuras N° 46 y 47 se obtiene que la probabilidad de incumplimiento pasa de un 28,3% en su nodo 0 a un 68,9% en su nodo 7; siendo la ruta entre estos nodos la regla requerida para los propósitos del proyecto, y ser utilizada en el proceso de la Sección 5.2.2.2.1 Planificar programas de fiscalización preventivos.

La regla o lógica de negocio para proponer la muestra de sostenedores y/o escuelas preventiva y focalizada a fiscalizar para el programa uso de recursos SEP responde a:

Si: ( $\text{Variación\_Matrícula} \leq -0,05$ ) and ( $\text{Tipo\_Apoyo\_Personal\_Propio} = 1$ ) and ( $\text{Matrícula\_2016} > 425$ ) and ( $\text{Planta} < 82$ ) and ( $\text{Retención} = 1$ ) implica que, la escuela tiene una probabilidad cercana al 70% de estar realizando usos indebidos de fondos SEP.

### **5.3.2 Asignación Óptima de Sostenedores y Escuelas**

Tal como fue descrito en la Sección 5.2.2.2.1 Planificar programas de fiscalización preventivos, en la actividad obtener muestra a fiscalizar para el programa cuenta con otra lógica de negocio que apoya la toma de decisiones al jefe de fiscalización en aquellos casos donde el tamaño de la muestra a fiscalizar sea superior a la cantidad de visitas disponibles y donde, el problema no sea simple de resolver. En estos casos, el analista de negocio es apoyado con el diseño e implementación de un algoritmo de optimización.

A continuación, se presenta el algoritmo de optimización que selecciona de entre un set de sostenedores y/o escuelas aquellas que generan el mayor beneficio, tras setear un conjunto de parámetros y atributos de los sostenedores y/o escuelas.

- Sets:

$j$  = escuelas (identificador único)

- Parámetros:

$v$  = cantidad de visitas asignadas a una zona

$e$  = cantidad de escuelas nunca fiscalizadas a incluir en una zona

$e_i$  = cantidad de escuelas particulares subvencionadas a incluir en una zona

$e_m$  = cantidad de escuelas municipales a incluir en una zona

$e_f$  = cantidad de escuelas fijas a incluir en una zona

$r_j$  = reintegro probable en MM\$ al fiscalizar la escuela  $j$

$n_j$  = 1 si la escuela  $j$  es del grupo nunca fiscalizadas, 0 si no

$p_j = 1$  si la escuela  $j$  es del grupo particulares subvencionadas, 0 si no

$m_j = 1$  si la escuela  $j$  es del grupo municipales, 0 si no

$f_j = 1$  si la escuela  $j$  es del grupo fijas, 0 si no

▪ Variables:

$e_j = 1$  si la escuela  $j$  incluida en la muestra a fiscalizar

$z =$  reintegro total en MM\$

▪ Restricciones:

La cantidad de escuelas a incluir en la muestra a fiscalizar, debe ser igual a la cantidad de visitas asignadas a una zona,

$$(1) \sum_{j \in e} e_j = v ;$$

La cantidad de escuelas del grupo nunca fiscalizadas a incluir en la muestra a fiscalizar, debe ser menor a la cantidad de escuelas nunca fiscalizadas asignadas a una zona,

$$(2) \sum_{j \in e} f_j \times e_j \leq e_i ;$$

La cantidad de escuelas del grupo particulares subvencionadas a incluir en la muestra a fiscalizar, debe ser menor a la cantidad de escuelas particulares subvencionadas asignadas a una zona,

$$(3) \sum_{j \in e} p_j \times e_j \leq e_i ;$$

La cantidad de escuelas del grupo municipales a incluir en la muestra a fiscalizar, debe ser menor a la cantidad de escuelas municipales asignadas a una zona,

$$(4) \sum_{j \in e} m_j \times e_j \leq e_i ;$$

La cantidad de escuelas del grupo fijas a incluir en la muestra a fiscalizar, debe ser menor a la cantidad de escuelas fijas asignadas a una zona,

$$(5) \sum_{j \in esc} f_j \times e_j \leq e_i ;$$

- Función Objetivo:

$$(5) M \quad (r_i \quad ) = z = \quad j r_j x e_j$$

El algoritmo parametrizable descrito se implementa con el software GAMS; por ejemplo, la Figura N° 48 ilustra las salidas de datos y resalta aquellas escuelas que conformarán la muestra a fiscalizar, tras configurar el algoritmo para que escoja las 10 escuelas que generen el mayor beneficio (detalles de la implementación ver Anexo N° 4).

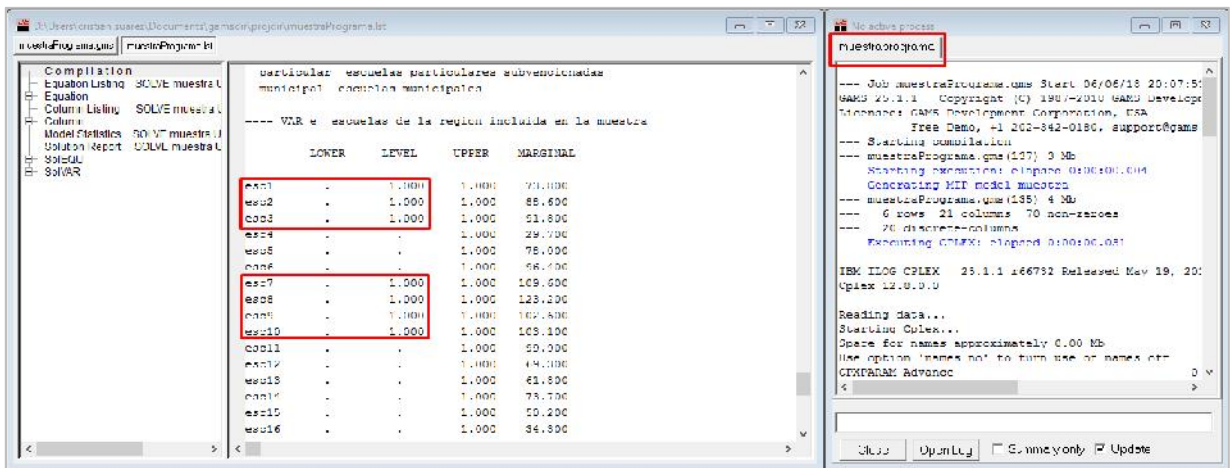


Figura N° 48: Resultados del algoritmo de optimización

### 5.3.3 Decidir Acciones a las Recomendaciones de Gastos SEP

Tal como fue descrito en la Sección 5.2.2.2.1 Planificar programas de fiscalización preventivos, y en particular en su actividad preparar recomendaciones automatizadas; esta lógica de negocio tiene por objeto definir las acciones a seguir en función del tipo de recomendación (tipo 1: uso indebido; tipo 2: error de imputación en cuenta de gasto; o tipo 3: escasa descripción del gasto) sugerida por un clasificador automático de textos, ideado para los propósitos del proyecto. Si la recomendación es del tipo 1, esta es incorporada como un ítem en cada visita de fiscalización preventiva (resultantes de la lógica anterior) a terreno para asesorar-educar al sostenedor y/o escuela, o efectuar el seguimiento y control de subsanación, al ser incluidas como una materia adicional en otros programas de fiscalización tradicionales (aprovechando la visita a terreno). Si es del tipo 2 o 3, esta es publicada en portal Web de rendición de cuentas.

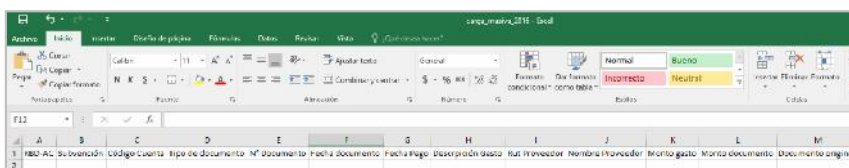
A continuación, se diseña el clasificador automático de textos instanciando las etapas descritas en la actividad elaborar modelos de clasificación de sostenedores y/o escuelas, pero con los ajustes requeridos por los modelos de text mining.

- Entendimiento del negocio:

Los sostenedores tienen la obligación de realizar la declaración anual de gastos SEP en los plazos establecidos por Ley; mientras que, la SUPEREDUC debe fiscalizar el debido uso de estos fondos e informar los incumplimientos detectados a los sostenedores en un plazo no superior a 6 meses, desde la fecha de cierre del proceso declarativo, de lo contrario toda acción posterior queda prescrita. Con esta forma de operar, la SUPEREDUC solicita reintegrar miles de millones de pesos a los sostenedores en incumplimiento, y tal como fue descrito estos fondos son de uso exclusivo y van en directo apoyo de los estudiantes vulnerables del país; por ello, es necesario, realizar acciones de prevención y/o mitigación para este tipo de incumplimiento.

- Entendimiento de los datos:

Los sostenedores rinden cuentas en un sistema TIC dispuesto para ello, cargando un archivo de texto (extensión .csv), ilustrado en la Figura N° 49, con la siguiente información: rbd/ac, identifica de manera única a cada escuela o sostenedor; subvención, señala el origen del recurso; código cuenta, imputación a la cual pertenece el gasto; tipo documento, documento financiero o comercial de respaldo; n° documento, identifica de manera única a cada documento; fecha documento, fecha de emisión del documento; fecha pago, fecha del desembolso de los recursos; descripción gasto, detalla el uso dado al bien o servicio adquirido; rut proveedor, identifica de manera única a cada proveedor; nombre proveedor, nombre de la persona natural o jurídica; monto documento, monto total del gasto; monto gasto, fracción del monto documento incurrido en el rbd/ac, subvención y código cuenta.



	A	B	C	D	E	F	G	H	I	J	K	L	M
1													
2													

Figura N° 49: Archivo de carga de gastos y sus atributos



Para diseñar el clasificador automático de textos, basta con seleccionar un código cuenta, al ser independientes unas de otras, y replicar el método a las demás. La cuenta seleccionada para la subvención = SEP es combustible y peaje (código cuenta = 410906<sup>22</sup>), que registra más de 41.100 imputaciones de gastos (26.314, es el promedio) y se ubica dentro de las 10 más utilizadas en la rendición de cuentas SEP.

- Preparación de los datos:

Previo a aplicar la analítica, tal como se ha descrito en el Capítulo 2: Marco Teórico, es necesario preprocesar los textos de interés; en este caso, el campo descripcion\_gasto ilustrado en la Figura N° 50 y donde los actuales sistemas presentan dificultades para procesarlo, integrarlo y convertirlo en información útil.

	A	B	C	D	E	F	G	H	I	J	K	L
1	pk_rendicion	fk_subvencion	fk_cuenta	RUB/AC	sostenedor	periodo	rut_documento	nombre_proveedor	fecha_documento	descripcion_gasto	monto_declarado	monto_total
47	16507821	1	410906	2885	50100500	2016	99520000-7	COMPANIA DE PETROLI	04-01-2016	BOJITA ELECTRONICA N° 226123 DE FECH	39999	39999
48	16789546	1	410906	24589	55945525	2016	76394171-K	ADMINISTRADORA DE E	12-12-2016	Acción: Fortalecer y promover la carativ	10000	10000
49	15557652	1	410906	384	69259500	2016	8832402-1	ARNOLD GUNTHER POT	28-09-2016	D.F.N.º 637 FONDO A RENDIR COMBUSTI	10000	30000
50	1008870	1	410906	2123	71914200	2016	9402000-7	INVERSIONES ENX S.A	10-11-2016	CARGA DE COMBUSTIBLE, REUNION EN S.	12009	12009
51	15555244	1	410906	384	69259500	2016	8653216-6	ARNOLDO ISAIAS TORR	28-09-2016	D.F.N.º 637 FONDO A RENDIR COMBUSTI	20000	30000
52	15568378	1	410906	2123	71914200	2016	96875230-8	RUTA MAIPO SOC. CON	15-11-2016	PEAJE DE RETORNO, REUNION EN SANTI	2300	2300
53	15568377	1	410906	2123	71914200	2016	96875230-8	RUTA MAIPO SOC. CON	15-11-2016	PEAJE DE IDA A REUNION EN SANTIAGO,	2300	2300
54	21730007	1	410906	387	69259500	2016	8653216-6	ARNOLDO ISAIAS TORR	25-08-2016	D.F.N.º 638 FONDO A RENDIR GASTOS DE	20000	20000
55	21200480	1	410906	20199	60114904	2016	40197349-2	Comercial y servicios te	09-11-2016	movilizacion compra coronas reposicion	20000	20000
56	17341580	1	410906	12545	70002810	2016	96765160-2	TERMINAL AGRICOLA SPH	09-11-2016	CANCELACION ESTACIONAMIENTO PARA	900	900
57	15561144	1	410906	2123	71914200	2016	76074265-6	OPERADORA DE ESTACI	10-05-2016	PAGO DE BENJINA, COMPRA DE VELAS PA	10000	10000
58	15700380	1	410906	2123	71914200	2016	76061338-K	BODEGAUE LOGISTICA Y	12-12-2016	RECARGA COMBUSTIBLE, VISITA EMPRES.	10000	10000

Figura N° 50: Extracto de datos<sup>23</sup>

El preprocesamiento a aplicar al campo descripcion\_gasto consiste en:

Tratar caracteres ilegibles, consiste en reemplazar aquellos códigos que no se encuentren en codificación UTF-8, utilizando tablas de conversión; por ejemplo, Ñ = Á; Ñ<sub>i</sub> = á; Ñ% = É; etc.

Tokenizar, para crear el vector de términos para cada texto, utilizando el api lucene para java (<http://lucene.apache.org/>).

Remover stopwords, extraer todos aquellos términos que no aportan contenido, utilizando una tabla de valores construida para este propósito (ver Tabla N° 15).

<sup>22</sup> La razón de presentar resultados en está cuenta y no en otra, se debe a privacidad solicitada por el área de negocio. Independiente de ello, la cuenta es representativa (documentos es superior al promedio).

<sup>23</sup> Los datos de los campos rbd/ac, sostenedor y rut\_documento, están enmascarados.

Tabla N° 15: Extracto tabla stopword y corrector ortográfico

Stopword		Corrector Ortográfico		
Id	Stopword	Id	Dice	Debe Decir
1	abel	1	abieta	abierta
2	abelar	2	academica	académica
3	abelardo	3	acensor	ascensor
4	accide	4	ach	achs
5	acosub	5	act.alumnos	alumnos
6	act	6	act.convivencia	convivencia
7	activ	7	act.cueca	cueca
8	actu	8	activdad	actividad
9	adb	9	activdaes	actividades
10	adios	10	actualizacion	actualización
11	adq	11	activdad	actividad

Aplicar corrector ortográfico, en paralelo con el punto anterior, se construye una tabla de datos con términos dice y debe decir (ver Tabla N° 15), para aplicar a cada vector de términos.

Identificar n-grams, tratar como un solo término aquellos que aportan valor juntos, se construye una tabla de datos con éstos.

Como resultado de los cinco pasos previos, se obtiene el primer diccionario para la subvención = SEP y código cuenta = 410906; este diccionario cuenta con 3.095 términos (notar que este trabajo es reutilizable para otras subvenciones y códigos de cuentas).

El preprocesamiento, para los propósitos del clasificador automático de textos, continúa asociando a cada término del diccionario sus variantes, sinónimos y/o formas verbales (ver Tabla N° 16); de forma tal, de construir redes de palabras por medio de una base de grafos<sup>24</sup> a la cual aplicar analítica.

---

<sup>24</sup> Se utiliza Neo4J versión 3.3.5

Tabla N° 16: Extracto de tablas de términos y sus variantes

Id	Nodo	Singular		Plural	
		Masculino-Neutro	Femenino	Masculino-Neutro	Femenino
1	alojamiento	alojamiento		alojamientos	
2	alternativo	alternativo	alternativa	alternativos	alternativas
3	alto	alto	alta	altos	altas
4	alumno	alumnado	alumnada	alumnos	alumnadas
5	alumno	alumno		alumnos	alumnas
6	avi	avi			
7	amarillo	amarillo	amarilla	amarillos	amarillas
8	ambiental	ambiental		ambientales	
9	ambiente	ambiente		ambientes	
10	amistoso	amistoso	amistosa	amistosos	amistosas
11	actividad	actividad			
12	acto	acto		actos	
13	actor	actor		actores	
14	actual	actual		actuales	
15	acuario	acuario		acuarios	

Id	Nodo	Gerundio	Participio Singular		Participio Plural	
			Masculino	Femenino	Masculino	Femenino
1	abrir	abriendo	abierto	abierta	abiertos	abiertas
2	accidental	accidentalando	accidental	accidentalada	accidentales	accidentales
3	accionar	accionando	accionado	accionada	accionados	accionadas
4	acercar	acercando	acercado	acercada	acercados	acercadas
5	acompañar	acompañando	acompañado	acompañada	acompañados	acompañadas
6	capacitar	capacitando	capacitado	capacitada	capacitados	capacitadas

Nodo: capacitar						
Presente	Preterito imperfe	Preterito perfe	Futuro	Condicional	Imperativo	
capacita	capacitaba	capacitó	capacitaré	capacitaría	capacita	
capacitas	capacitabas	capacitaste	capacitarás	capacitarías	capacite	
capacita	capacitaba	capacitó	capacitará	capacitaría	capacitemos	
capacitamos	capacitabamos	capacitamos	capacitaremos	capacitaríamos	capacitad	
capacitáis	capacitabais	capacitasteis	capacitaréis	capacitaríais	capaciten	
capacitan	capacitaban	capacitaron	capacitarán	capacitarían		

Previo a construir la base de términos de grafos dirigidos, por convención se utiliza:

- (1) Término tipo verbo, el nodo padre o raíz es el infinitivo al cual se asocian todas las formas verbales (y no entre ellas).
- (2) Otros tipos, el nodo padre o raíz es el lema (si aplica) o su forma neutra al cual se asocian todas las variantes (y no entre ellas).
- (3) Las relaciones entre términos sinónimos, se ligan entre nodos padres.

Esta convención, permite simplificar las redes semánticas y aumenta las posibilidades de acceder de un nodo a otro, ya sea directamente si la relación está explícita o indirectamente a través de sus variantes, tal como ilustra la Figura N° 51.

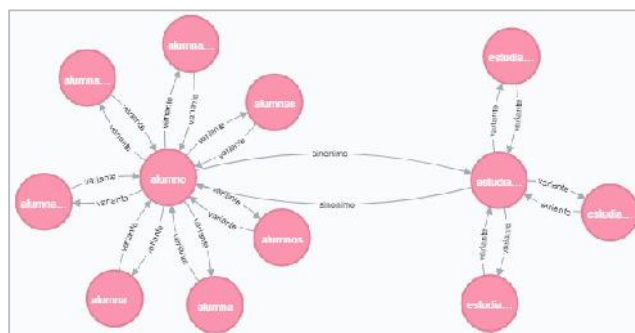


Figura N° 51: Extracto red semántica de grafos

La Figura N° 51 ilustra el resultado del proceso previo y cargado en un motor de base de grafos, que cuenta con más de 6.000 términos y más de 10.000 relaciones.

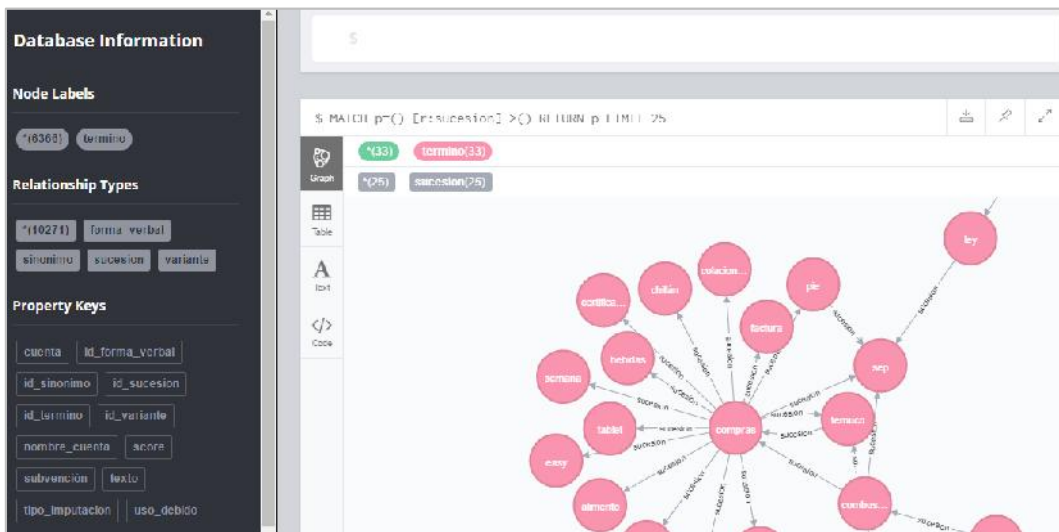


Figura N° 52: Red semántica en motor de base de grafos

- Modelamiento:

Para apoyar el preprocesamiento de los más de 41.000 textos del campo descripción\_gasto en la subvención = SEP y código cuenta = 410906, y cargarlos a una base de grafos, como parte del alcance del proyecto, se implementa una solución TIC ad hoc que es descrita en detalle en el Capítulo 6: Propuesta de Apoyo Tecnológico.

En la Figura N° 53 se ilustra la relación entre las etapas de procesamiento de textos, el clasificador y los textos clasificados.

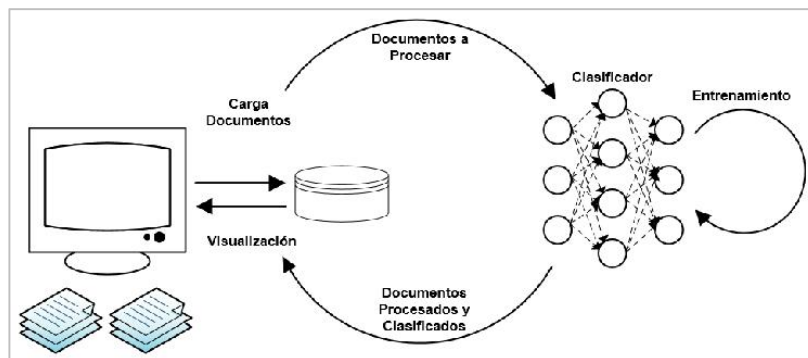


Figura N° 53: Etapas de la clasificación automática de textos

Clasificación manual, tal como se ha descrito, todo método supervisado requiere de una etapa de clasificación manual; por ello, se seleccionan aleatoriamente 1.000 vectores de términos (aprox. 2,5% del total) para que un experto de negocio en la SEP los clasifique

según su uso en debidos o indebidos; estos vectores clasificados son marcados en la relación de sucesión de la red semántica (ver Figura N° 54).

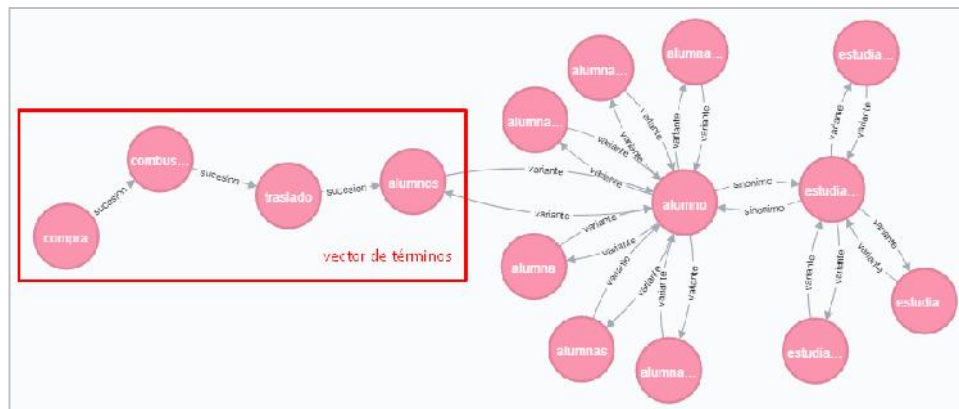


Figura N° 54: Extracto red semántica uso debido, SEP y código cuenta = 410906

Luego otro grupo de vectores de términos, es clasificado por el modelo y sus resultados son revisados por el experto (quien visa o corrige la clasificación) y se cargan a la red semántica; en un tercer grupo de vectores de términos, se obtienen resultados con un recall (0,8) y precisión (0,67) y un rendimiento del método superior al 70%, considerado estos resultados como aceptables.

Ponderadores de nodos y relaciones de sucesión, tras culminar la etapa de entrenamiento y para la red de uso indebido para la subvención = SEP y código cuenta = 410906, se calcula la frecuencia de aparición e inversa (método TF-IDF) obteniendo los pesos de cada término en el vector de términos, y con ello, descubrir términos relevantes; información útil que favorecerá el aumento de precisión del modelo.

Clasificación automática, utiliza el método descrito en el Capítulo 2: Marco Teórico, donde: un vector de palabras  $h$  que no ha sido clasificado, se reconstruye en la red semántica de la categoría (en este caso subvención = SEP, código cuenta = 410906 y tipo uso = Indebido); y donde, si el costo de reconstrucción es bajo, con alta probabilidad  $h$  pertenece a la categoría (en este caso, es un uso indebido).

Para calcular el costo de reconstrucción  $C = \left(\frac{l^\alpha}{p^\beta}\right) + v$ , se implementa una clase java que utiliza el api de Neo4j para realizar consultas en Cypher Query Language a la base de grafos (red semántica) para obtener los valores requeridos por  $C$ .

Para explicitar el cálculo del costo de reconstrucción, se recurre al siguiente ejemplo de descripcion\_gasto: compra de combustible para calefacción de la escuela y sus estudiantes; en este caso,  $h = [compra, combustible, calefacción, escuela, estudiantes]$  el vector de términos a clasificar.

El vector anterior, es consultado en la red semántica de uso indebido para obtener el grafo de rutas más cortas entre cada par de términos, usando la siguiente sentencia:

```
MATCH (termino_1:termino { texto: 'compra' }),(termino_2:termino { texto: 'combustible'
}), p = shortestPath((termino_1)-[*..6]->(termino_2)) RETURN p UNION MATCH
(termino_1:termino { texto: 'combustible' }),(termino_2:termino { texto: 'calefacción' }), p =
shortestPath((termino_1)-[*..6]->(termino_2)) RETURN p UNION MATCH
(termino_1:termino { texto: 'calefacción' }),(termino_2:termino { texto: 'escuela' }),
p = shortestPath((termino_1)-[*..6]->(termino_2)) RETURN p UNION MATCH
(termino_1:termino { texto: 'escuela' }),(termino_2:termino { texto: 'estudiantes' }), p =
shortestPath((termino_1)-[*..6]->(termino_2)) RETURN p
```

El resultado de ejecutar la sentencia cypher, se ilustra en Figura N° 55.

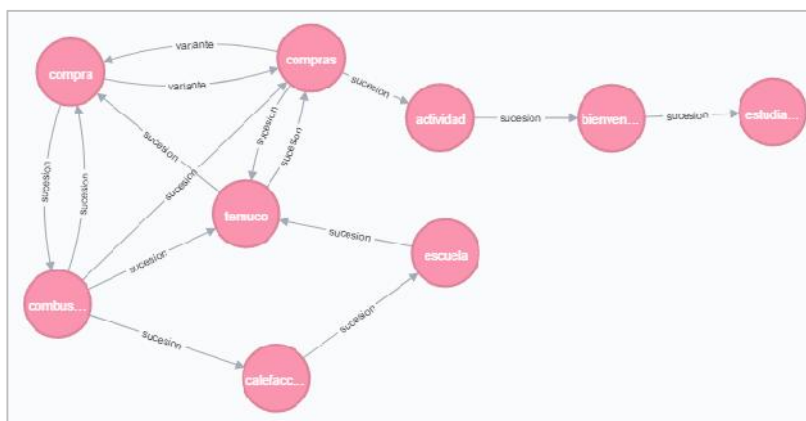


Figura N° 55: Grafo de rutas más cortas para el vector de términos

Siendo el algoritmo que calcula el costo de reconstrucción del vector  $h$  es:

- 1) Posicionarse en termino  $i$  ( $1 \leq i \leq \text{cardinalidad de } h$ ); en este caso, "compra".
- 2) Encontrar la ruta más corta para llegar al término  $i + 1$ ; combustible.
  - a. Contar cantidad relaciones de tipo sucesión ( $s$ ); en este caso, 1.

- b. Contar cantidad relaciones de tipo variante ( $v$ ); en este caso, 0.
  - c. Contar cantidad relaciones de tipo forma verbal ( $fv$ ); en este caso, 0.
- 3) Cálculo de  $l(i) = s$  . Las relaciones  $v$  y  $fv$  tiene costo 0; en este caso,  $l = 1$ .
  - 4) Cálculo de  $l = l + l(i)$ ; donde,  $l$  se inicia en 0 en primera iteración.
  - 5) Iterar y volver al punto 1); el valor de  $i = i + 1$ ; hasta que  $i <$  cardinalidad de  $h$ .
  - 6) Cálculo de  $P =$  cantidad de términos  $h$  presentes en el grafo.
  - 7) El valor de  $\alpha = 2$  y  $\beta = 1$ , corresponden a las ponderaciones de  $l$  y  $P$ .
  - 8) El valor de  $C^* = (1 + 1 + 1 + 25) / 5 = 5,6$ .

Para clasificar si el vector  $h$  se reconstruye en el grafo  $C^*$  se compara contra el costo mínimo y máximo de  $C$  para el vector  $h$ ;  $C_{min}$  se obtiene, si todos los pares de nodos se conectan en una relación de tipo sucesión=  $4/5 = 0,8$ ;  $C_{max}$  se obtiene si todos los pares de nodos se conectan en su tope de 6 relaciones de tipo sucesión =  $36 \times 5/5 = 36$ .

Con lo anterior, se calcula  $C$  normalizado =  $C^* \times 100 / (C_{max} - C_{min}) = 15\%$ . Ahora el clasificador automático de textos, en función del valor de corte de  $C$  normalizado seteado en 30%, señala que si es menor a 30% el vector  $h$  se reconstruye a bajo costo implicando que el vector  $h$  pertenece a la categoría y en este caso, es de uso indebido.

Respecto al valor de  $v$ , términos relevantes presentes en  $h$ , se utiliza de forma discriminante para la categoría; es decir,  $v$  es mayor que 0; el vector  $h$  corresponde a un uso indebido.

La siguiente tabla de datos ilustra los resultados del clasificar, donde se aprecia los tipos de recomendación tipo 1: uso indebido de recursos SEP (solicita instruir al proceso de gestión de las fiscalizaciones incorporarla como una materia en cada visita de fiscalización a terreno con objeto de fiscalizar-educar al sostenedor y/o escuela que se encuentre en incumplimiento), tipo 2: error de imputación en cuenta de gasto o tipo 3: escasa descripción del gasto (solicita publicar en el sistema de rendición de cuentas).

Tabla N° 17: Extracto de archivo de salida del clasificador automático de textos

Línea	Vector Preprocesado	Semántica	Nodos en Grafo	Relaciones Sucesión	Discriminantes	Costo Reconstrucción	Monto Declarado	Monto Total	Recomendación
34	petróleo	Semántica débil	0	0	0	1000000	38859	38859	Tipo 3
30	petróleo	Semántica débil	0	0	0	1000000	20000	20000	Tipo 3
38	puajo	Semántica débil	0	0	0	1000000	3870	3870	Tipo 3
1447	giro destinado gastos alojamiento salida pedagogico valpara	Semántica normal	10	0	0	0,18	16000	16000	Tipo 2
2231	lirila	Semántica débil	0	0	0	1000000	126081	126081	Tipo 2
2319	para entrada mano feria municipal tola	Semántica normal	0	0	0	1000000	1000	1000	Tipo 1
2441	cuadernos blocks pinceles carpetas lápices	Semántica normal	5	4	0	0,16	1702390	1702390	Tipo 2
2537	microsoft office	Semántica normal	2	1	0	0,25	99980	99980	Tipo 2
2746	insumos computacionales	Semántica normal	2	1	0	0,25	271780	271780	Tipo 2
3302	eslacionamiento	Semántica débil	0	0	0	1000000	1800	1800	Tipo 2
4174	accesorios presentación banda patrias	Semántica normal	0	0	0	1000000	2100	2100	Tipo 2
4870	pellet soporte plasma	Semántica normal	0	0	0	1000000	39420	39420	Tipo 2
4908	copio llaves	Semántica normal	0	0	0	1000000	3000	3000	Tipo 2
4958	aparcamiento	Semántica débil	0	0	0	1000000	1500	1500	Tipo 2
5251	descripción	Semántica débil	0	0	0	1000000	10000	10000	Tipo 3
5296	extra	Semántica débil	0	0	0	1000000	28617	28617	Tipo 1
5377	lubricantes	Semántica débil	0	0	0	1000000	33500	33500	Tipo 2
5545	filter	Semántica débil	0	0	0	1000000	11800	11800	Tipo 2
6282	Nulo	Semántica nula	0	0	0	1000000	10000	10000	Tipo 1
6606	encuadrado documentos	Semántica normal	0	0	0	1000000	4800	4800	Tipo 2
6827	calefactor bosca	Semántica normal	0	0	0	1000000	274800	274800	Tipo 1
7330	compra chapitas ceremonia lectura	Semántica normal	0	0	0	1000000	5000	5000	Tipo 2

▪ Evaluación:

La siguiente tabla describe los tipos de incumplimientos SEP, en promedio en los últimos años, y de la cual se desprende que los resultados obtenidos apuntar a prevenir y/o mitigar, por medio de recomendaciones, el tipo: gasto no se ajusta al objeto de la subvención o aporte que corresponde a más del 60% de los gastos no aceptados.

Tabla N° 18: Incumplimientos SEP promedio últimos años

Tipo de Incumplimiento	Gastos no Aceptados	
	Promedio	Porcentaje
Gasto no se ajusta al objeto de la subvención o aporte	9.817.314.870	60,6%
Documentación enmendada o incorrecta	2.989.831.524	18,5%
Gasto rendido o pagado en período incorrecto	1.150.687.998	7,1%
Sin documentación de respaldo	877.108.004	5,4%
Documentación de respaldo incompleta	529.023.160	3,3%
Gasto no vinculado a las acciones del PME	255.064.220	1,6%
Producto o servicio no recibido	195.590.265	1,2%
Otros	375.144.875	2,3%
<b>Total</b>	<b>16.189.764.914</b>	<b>100,0%</b>



El clasificador para la cuenta de gasto combustible y peajes logra los siguientes resultados (resumen de la tabla N°17).

Tabla N° 19: Recomendaciones cuenta combustible y peaje recursos año 2016

Tipo de Recomendación	Documentos	% Documentos	Monto MM\$	% Monto MM\$
tipo 1: uso indebido de recursos SEP	464	1,1%	26,4	1,7%
tipo 2: error de imputación en cuenta de gasto	3.653	8,9%	86,2	5,5%
tipo 3: escasa descripción del gasto	26.860	65,2%	982,8	62,7%

Los cuales, al ser comparados con los resultados obtenidos en años previos, para la misma cuenta de gasto, con los métodos tradicionales se presentan en la siguiente tabla.

Tabla N° 20: Comparación resultados años previos versus clasificador

Año 2014		Año 2015		Clasificador	
Documentos	Monto MM\$	Documentos	Monto MM\$	Documentos	Monto MM\$
312	134	1186	86	30.977	1.095

De la tabla anterior, se desprende la gran capacidad de detección de documentos riesgosos que cuenta el clasificador automático basado en redes de palabras, al trabajar con la información semi – estructurada de las rendiciones de cuentas y donde los actuales sistemas de la SUPEREDUC presentan dificultades para procesarla y convertirla en información útil.

## CAPÍTULO 6: PROPUESTA DE APOYO TECNOLÓGICO

Para elaborar la propuesta de apoyo tecnológico se utilizan las directrices dadas por Jacobson, Booch & Rumbaugh (2000) para realizar el relevamiento de requisitos, casos de uso y arquitectura; y de Larman (2010) respecto de los patrones de diseño de software.

El apoyo tecnológico propuesto para el proyecto tiene por alcance apoyar al proceso analizar comportamientos de sostenedores y escuelas, para preparar fiscalizaciones preventivas y focalizadas para el programa uso de recursos SEP, y preparar recomendaciones oportunas a la declaración anual de gastos; al proceso planificar programas de fiscalización preventivos, para obtener la muestra a fiscalizar preventiva y focalizada para el programa en cuestión y automatizar las acciones de prevención y/o mitigación asociadas a las recomendaciones generadas; y a los procesos de gestión y ejecución de las fiscalizaciones, para efectuar seguimiento y control de sus actividades. En resumen, el apoyo tecnológico propuesto para el programa de fiscalización uso de recursos SEP, es descrito en los siguientes módulos: fiscalización preventiva y focalizada; recomendaciones automatizadas; y seguimiento y control de visitas a terreno.

A continuación, se describe el tipo de apoyo tecnológico requerido para cada módulo.

- Módulo fiscalización preventiva y focalizada:

Tal como describe la Sección 5.2.2 Diseño en BPMN el apoyo tecnológico requerido corresponde a un sistema especializado en data mining y optimización, el cual se compone por un servidor de datos y de aplicaciones; en el servidor de datos se instala el *datawarehouse* (ideado para el proyecto, luego puede ser migrado) mientras que, en el servidor de aplicaciones se instalan las licencias de los softwares SPSS Modeler y GAMS. El otro tipo de apoyo tecnológico requerido para este módulo consiste en el desarrollo de una pieza de software que permita registrar de manera ordenada y/o en fichas la información obtenida de los modelos analíticos para el programa de fiscalización en cuestión y el ranking en caso programas de fiscalización tradicional.

- Módulo recomendaciones automatizadas:

Tal como describe la Sección 5.2.2 Diseño en BPMN el apoyo tecnológico requerido corresponde a una pieza de software donde, su primer componente extrae, prepara y

almacena en formato ad hoc la información diaria de las declaraciones anuales de gastos; otro componente ejecuta el clasificador automático de textos de descripciones de gastos y crea el archivo de texto con listado de recomendaciones; finalmente, otro componente lee el archivo anterior y función del tipo de recomendación, decide la acción de prevención y/o mitigación a aplicar a cada caso (enviarla al proceso de gestión de las fiscalizaciones o publicarla en el portal Web de rendición de cuentas).

- Módulo seguimiento y control visitas a terreno:

El tipo de apoyo tecnológico requerido corresponde a la implementación de un *datamart* e indicadores de seguimiento y control (cubos OLAP) utilizando la información transaccional del actual sistema TIC que soporta los procesos de gestión y ejecución de las fiscalizaciones (SIFE).

## 6.1 ESPECIFICACIÓN DE REQUERIMIENTOS

### 6.1.1 Requerimientos Funcionales: Fichas de los Programas

Req\_01: la autenticación del sistema solicita nombre usuario, contraseña y perfil. El nombre usuario y contraseña deben utilizar el protocolo LDAP<sup>25</sup> (*Lightweight Directory Access Protocol*) institucional y el perfil es seleccionable del siguiente listado: jefe unidad datos, responsable programa, analista y administrador.

Req\_02: una vez autenticado un usuario, el sistema autoriza y carga sólo las transacciones (con los permisos: crear, actualizar, leer, eliminar) habilitadas para el perfil seleccionado (ver Tabla N° 21).

Req\_03: el sistema debe permitir al perfil jefe unidad de datos, crear un nuevo programa de fiscalización y tras ello, permitir editar los siguientes campos: nombre (texto largo 100), descripción (texto largo 500), seleccionar tipo programa (recurso o normativo), año de ejecución (numérico mayor a 2015 y menor a 2100), seleccionar el responsable del

---

<sup>25</sup> Protocolo de acceso unificado a un conjunto de información sobre una red, entre ellas, usuario y contraseña del personal interno.

programa (listado de usuarios con ese perfil), vigencia: fecha inicio (fecha mayor 01-01-2015 y menor 31-12-2100) y vigencia: fecha fin (fecha mayor que fecha inicio).

Req\_04: el sistema previo a registrar un nuevo programa debe validar que todos los datos ingresados (en Req\_03) sean distintos de nulos o vacíos. Además, valida que la tupla: nombre, tipo programa, año de ejecución; sea única; si la tupla no existe, se persiste la información en la base de datos, creando el identificador único (numérico autocorrelativo) y con estado en valor vigente (1 = vigente, 0 = no vigente).

Req\_05: el sistema debe permitir al perfil jefe unidad datos una vez creado un nuevo programa de fiscalización, agregar información adicional, asociar o eliminar obligaciones normativas (desde un listado precargado), registrar o eliminar cada uno de sus objetivos (texto largo 500) y adjuntar en formato digital (.pdf o .doc tamaño máximo 5Mb) información adicional.

Req\_06: el sistema debe permitir al perfil responsable programa una vez seleccionado un programa de fiscalización vigente, editar y visar la información ingresada.

Req\_07: el sistema debe permitir al perfil responsable programa una vez seleccionado un programa de fiscalización vigente, ingresar manualmente o por carga masiva (.csv) los sostenedores y/o escuelas (sujetos de fiscalización) y el clúster de riesgo al que pertenecen para el programa (clúster 1, ..., clúster n; donde 1 más riesgoso).

Req\_08: el sistema debe permitir al perfil analista una vez seleccionado un programa de fiscalización vigente, crear un nuevo criterio de riesgo y con ello, editar los siguientes campos: nombre criterio (texto largo 150), descripción (texto largo 500), seleccionar ámbito (documento, sostenedor o escuela) y peso (numérico entre 0 y 1 con tres cifras significativas, formato x, xx); además, la suma de todos los pesos debe ser 1.

Req\_09: el sistema debe permitir al perfil analista una vez seleccionado un programa de fiscalización y un criterio de riesgo vigente, ingresar por carga masiva (.csv) los resultados obtenidos para el criterio; siendo las columnas del archivo.csv: id (numérico que representa el identificador único del documento rendido); rut/rbd (numérico que representa el identificador único de sostenedor o escuela); ámbito (documento, sostenedor o escuela); probabilidad (número entre 0 y 1 con tres cifras significativas,

formato x, xx); impacto (número entero mayor que cero), que en caso del programa uso de recursos SEP, corresponde a la suma de los montos de los documentos riesgosos.

Req\_10: el sistema debe permitir al perfil analista una vez seleccionado un programa de fiscalización y cargados todos los criterios de riesgo vigentes, calcular automáticamente el ranking para el programa (numérico mayor que cero con tres cifras significativas, formato x, xx) por sostenedor y escuela; donde, el ranking es la sumatoria de probabilidades por sus impactos.

Req\_11: el sistema debe permitir al perfil administrador, gestionar usuarios, perfiles y autorizaciones de cada perfil; en función de la siguiente tabla.

Tabla N° 21: Perfiles y autorizaciones

	Jefe Unidad Datos	responsable Programa	Analista	Administrador
Nuevo Programa de Fiscalización	C, A, L, E	A, L	L	
Información Adicional al Programa de Fiscalización	C, A, L, E	C, A, L, E	L	
Sujetos de Fiscalización para el Programa de Fiscalización	L	C, A, L, E	L	
Clúster de Riesgo para el Programa de Fiscalización	L	C, A, L, E	L	
Criterios de Riesgo	A, I	I	C, A, I, F	
Resultados Criterios de Riesgo	L	L	C, A, L, E	
Ranking del Programa de Fiscalización	L	L	C, A, L, E	
Usuarios - Perfiles				C, A, L, E

C = Crear; A = Actualizar; L = Leer; E = Eliminar

### 6.1.2 Requerimientos Funcionales: Recomendaciones

Req\_01: la autenticación del sistema solicita nombre usuario, contraseña y perfil. El nombre usuario y contraseña deben utilizar el protocolo LDAP institucional y el perfil es seleccionable del listado: analista negocio, experto negocio y administrador.

Req\_02: el sistema de conectarse automáticamente y una vez por día a la base de datos del sistema de rendición de cuentas para extraer la información cargada por los sostenedores en las distintas cuentas de gastos; por cada cuenta de gasto genera un archivo de texto (.csv) con los campos: identificador del gasto (numérico), sostenedor (numérico), rbd (numérico), descripción del gasto (texto largo 500), cuenta de imputación (numérico), rut proveedor (numérico) y monto documento (numérico).

Req\_03: el sistema debe permitir al perfil analista negocio, cargar un archivo de texto (.csv) con los campos identificador del gasto (numérico) y descripción del gasto (texto largo 500); tras seleccionar la subvención del listado: escolaridad, SEP, mantenimiento, PIE, prortención, internado, refuerzo educativo y FAEP; y la cuenta imputación del

listado: transporte escolar, materiales de oficina, reproducción de documentos, alimentación, insumos computacionales, combustible y peaje, materiales, entre otros.

Req\_04: el sistema debe almacenar cada archivo de texto (.csv) cargado y cada registro (fila del archivo) debe ser insertado en una tabla de datos con los siguientes campos: identificador del gasto (numérico, mayor que cero y autoincremental), descripción del gasto (texto largo 500), subvención y código cuenta (ambos de los listados antes descritos) y procesado (boolean verdadero o falso).

Req\_05: el sistema debe automáticamente por cada fila de la tabla anterior donde el campo procesado sea falso, tokeniza el campo descripción del gasto y crea el respectivo vector de términos. A cada vector de términos, se remueven automáticamente los stopword, se aplica el corrector ortográfico y se generan los n - grams.

Req\_06: el sistema debe permitir al perfil analista negocio, acceder a un reporte Web con todos los vectores de términos procesados y/o para descarga en un archivo de texto (.csv), para que el perfil analista negocio identifique nuevos términos: stopword, correctores ortográficos y n - grams.

Req\_07: el sistema debe permitir al perfil analista negocio, mantener (individual o por carga masiva) la tabla de datos stopword con los siguientes campos: identificador del stopword (numérico, mayor que cero y autoincremental), término (texto largo 100), estado (boolean verdadero o falso) y fecha creación (fecha y hora).

Req\_08: el sistema debe permitir al perfil analista negocio, mantener (individual o por carga masiva) la tabla de datos revisor ortográfico con los siguientes campos: identificador del revisor (numérico, mayor que cero y autoincremental), término dice y termino debe decir (ambos, texto largo 100), estado (boolean verdadero o falso) y fecha creación (fecha y hora).

Req\_09: el sistema debe permitir al perfil analista negocio, mantener (individual o por carga masiva) la tabla de datos n-grams con los siguientes campos: identificador del n-grams (numérico, mayor que cero y autoincremental), cadena de términos (texto largo 100), estado (boolean verdadero o falso) y fecha creación (fecha y hora).

Req\_10: el sistema debe permitir al perfil analista negocio, mantener (individual o por carga masiva) la tabla de datos nodos términos con los siguientes campos: identificador del término (numérico, mayor que cero y autoincremental), término (texto largo 100), estado (boolean verdadero o falso) y fecha creación (fecha y hora). Además, el sistema automáticamente tras cada mantención, debe actualizar una base de grafos.

Req\_11: el sistema debe permitir al perfil analista negocio, mantener (individual o por carga masiva) la tabla de datos relación sucesión con los siguientes campos: identificador de la relación (numérico, mayor que cero y autoincremental), término uno y termino dos (ambos, texto largo 100), subvención y código cuenta (ambos de los listados antes descritos), tipo uso (debido o indebido), estado (boolean verdadero o falso) y fecha creación (fecha y hora). Además, el sistema automáticamente tras cada mantención, debe mantener la base de grafos.

Req\_12: el sistema debe permitir al perfil experto negocio, entrenar la red semántica de una determinada subvención, código cuenta y clasificador tipo uso (ambos de los listados antes descritos). Al cargar un archivo de texto (.csv) con todas aquellas descripciones de gastos procesadas, el sistema automáticamente debe crear las relaciones de sucesión en tablas de datos y base de grafos, y aumentar en una unidad cada ocurrencia en la tabla nodo términos y relación sucesión en el campo score (numérico).

Req\_13: el sistema debe permitir al perfil experto negocio, una vez entrenada la red semántica de una determinada subvención, código cuenta y clasificador tipo uso (ambos de los listados antes descritos) calcular el TF - IDF para apoyar en la identificación de términos relevantes; además, el sistema debe permitir a este perfil, marcar los términos relevantes o discriminantes para la categoría seleccionada.

Req\_14: el sistema debe permitir al perfil analista negocio, tras seleccionar una determinada subvención, código cuenta y clasificador tipo uso (ambos de los listados antes descritos), clasificar cada descripción de gasto cargada (individual o masiva) según algoritmo descrito en la lógica de negocio y tipos de recomendación; donde, los resultados se disponen en un reporte Web y/o en un archivo de texto (.csv).

Req\_15: el sistema debe cargar automáticamente en la tabla de recomendaciones de la base de datos del sistema de rendición de cuentas, las recomendaciones de tipo 2 y 3.

### **6.1.3 Requerimientos No Funcionales**

Req\_01: toda funcionalidad del sistema, salvo la ejecución del clasificador automático de textos, debe responder al usuario en menos de 5 segundos.

Req\_02: tras modificar datos, la actualización de la base de datos del sistema, debe disponerse a los perfiles autorizados para acceder a ella, en menos de 2 segundos.

Req\_03: el sistema debe ser desarrollado en base a la arquitectura técnica y los patrones de desarrollo de software definidos por la SUPEREDUC.

Req\_04: el sistema, software y base de datos, debe ser respaldado semanalmente.

Req\_05: el sistema debe proporcionar mensajes de error claros y orientados al usuario.

Req\_06: el sistema web debe poseer un diseño *responsive*, para ser visualizada en computadores personales, tablet y teléfonos inteligentes.

Req\_07: el sistema debe tener una disponibilidad del 99,98%.

## **6.2 ARQUITECTURA TECNOLÓGICA**

La arquitectura tecnológica requerida por el proyecto se basa en las directrices de Jacobson, Booch & Rumbaugh (2000); los cuales, señalan que los directores de la arquitectura son los casos de uso y otros factores como: experiencia en trabajos previos, patrones arquitectónicos, productos software requeridos (sistemas operativos y base de datos), middleware requerido, interoperabilidad con sistemas heredados, estándares y políticas corporativas a adoptar, requisitos no funcionales y distribución del sistema.

### **6.2.1 Arquitectura de Referencia**

El proyecto instancia la arquitectura de referencia establecida por la SUPEREDUC, ilustrada en la Figura N° 56; donde, la solución tecnológica atiende sólo clientes internos, los que interactúan por medio de un *web browser* que se autentican contra el componente LDAP (Microsoft). La capa intermedia de la solución tecnológica, se resuelve por medio de procesos masivos *integration services* (Microsoft), piezas de software desarrolladas a la medida (lenguaje de programación java 7, Oracle), analítica basada en datos (modelos segmentación y clasificación, SPSS Modeler) y reglas de optimización (programación



lineal, Gams). La capa de datos de la solución tecnológica, utiliza el *datawarehouse* y *datamart* en SQLServer 2016 (Microsoft), bases de datos transaccionales (PostgreSQL 2015), cubos OLAP en *analysis services* (Microsoft) y bases de grafos (Neo4j).

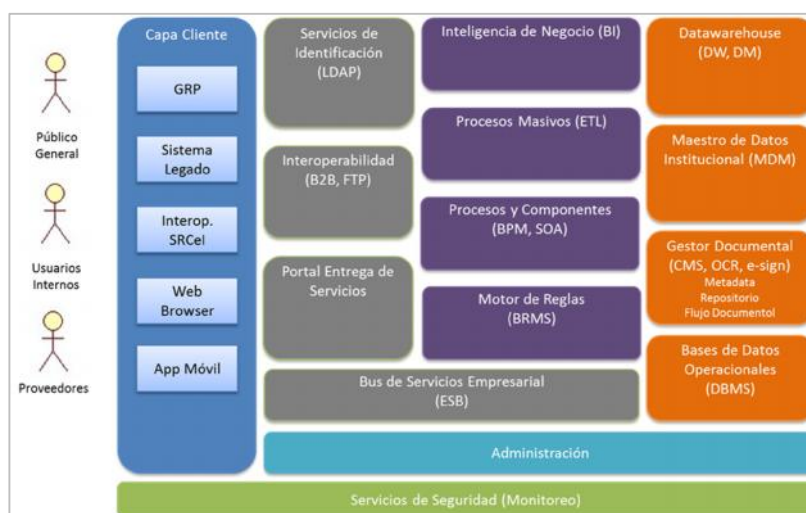


Figura N° 56: Arquitectura de referencia SUPEREDUC

### 6.2.2 Framework y Patrones de Diseño

Framework, define, en términos generales, un conjunto estandarizado de conceptos, prácticas y criterios para enfocar una problemática en particular, y sirve de referencia para enfrentar y resolver nuevos problemas de índole similar.

Patrones de diseño, son la base para la búsqueda de soluciones comunes en el desarrollo de software y otros ámbitos referentes al diseño de interacción o interfaces; para ser considerado un patrón, una solución debe haber comprobado su efectividad resolviendo problemas similares en ocasiones anteriores y debe ser reutilizable.

La solución tecnológica propuesta por el proyecto, utiliza los siguientes framework<sup>26</sup>: java server faces, facelet, primefaces, mybatis y log4j; junto con, los siguientes patrones de diseño: modelo vista controlador, data access object y transfer object.

<sup>26</sup> Detalles de los Framework y Patrones de Diseño en Anexo N° 3.

### 6.2.3 Diagrama de Despliegue y Componentes

La solución tecnológica propuesta por el proyecto, es desplegado en una máquina con sistema operativo Windows Server que tendrá instalada y configurada un servidor de aplicaciones JBOSS, el cual brinda un buen rendimiento a las tecnologías Java EE 7+; junto con, tener instalada y configurada una base de datos relacional PostgreSQL 8.4 y una base de grafos Neo4J 3.3.5.

El siguiente diagrama ilustra los componentes desplegados físicamente y el modo de comunicación entre ellos.

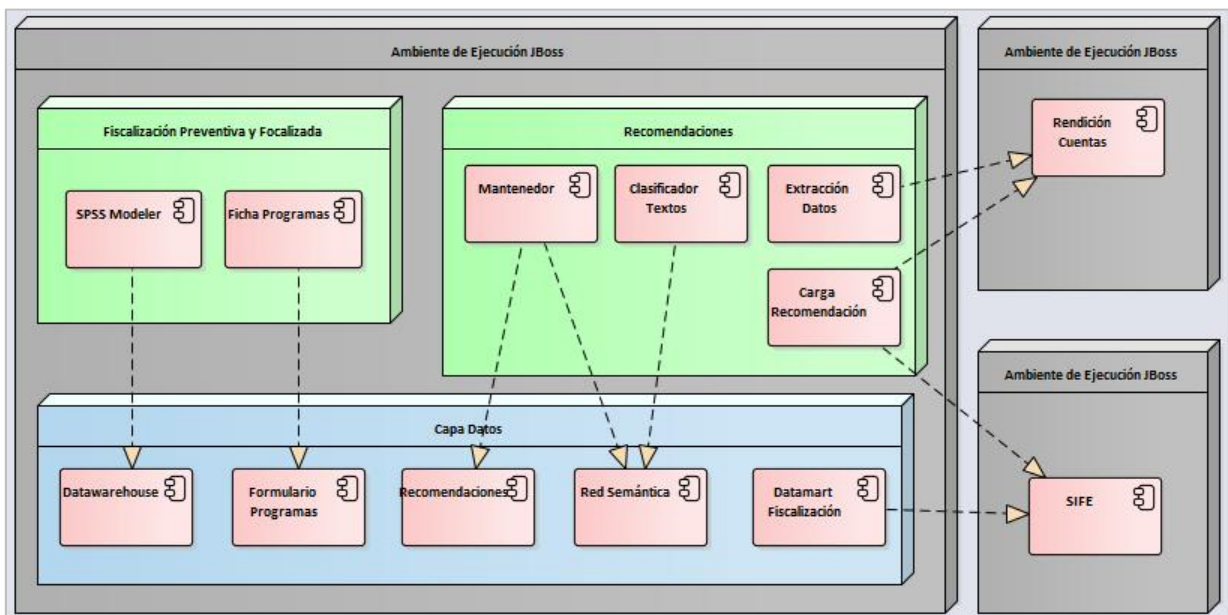


Figura N° 57: Diagrama de despliegue y componentes

## 6.3 DISEÑO DE LA APLICACIÓN

### 6.3.1 Casos de Uso: Fichas de los Programas

Caso Uso 01	Login
Objetivo	Permitir el acceso al sistema a aquellos perfiles y usuarios autorizados.

Actores	Jefe Unidad de Datos, Analista, Responsable del Programa y Experto Negocio
Precondiciones: el usuario debe estar registrado en LDAP institucional y creado en el sistema con alguno de los perfiles posibles.	
Breve descripción: el actor debe ingresar su nombre de usuario, clave LDAP y perfil; y el sistema solo permite acceso si la tupla está autorizada, en caso contrario informa al usuario la situación.	

Caso Uso 02	Crear/editar ficha programa de fiscalización
Objetivo	Mantener la información de cada programa de fiscalización clasificada y centralizada.
Actores	Jefe Unidad de Datos y Responsable del Programa
Precondiciones: el usuario debe estar autenticado en el perfil requerido.	
Breve descripción: los perfiles autorizados crean y mantienen la información de cada programa de fiscalización; para ello, deben registrar el nombre del programa, una breve descripción, tipo programa (recurso o normativo) y vigencia o fechas de operación.	

Caso Uso 03	Asociar obligaciones normativas
Objetivo	Registrar las obligaciones normativas que debe cautelar el programa de fiscalización, información necesaria para desarrollar los criterios de riesgo y/o modelos analíticos.
Actores	Jefe Unidad de Datos y Responsable del Programa
Precondiciones: el usuario debe estar autenticado en el perfil requerido y tener un programa de fiscalización vigente.	

Breve descripción: los perfiles autorizados buscan en un listado precargado las obligaciones normativas vigentes (una o más) y las asocia (o quita) al programa de fiscalización.

Caso Uso 04	Asociar registrar objetivos
Objetivo	Registrar los objetivos, metas e indicadores que debe cumplir el programa de fiscalización, información necesaria para desarrollar los criterios de riesgo y/o modelos analíticos, y evaluar los resultados tras ejecución.
Actores	Jefe Unidad de Datos y Responsable del Programa
Precondiciones: el usuario debe estar autenticado en el perfil requerido y tener un programa de fiscalización vigente.	
Breve descripción: los perfiles autorizados redactan (o editan) cada uno de los objetivos, metas e indicadores a cumplir.	

Caso Uso 05	Adjuntar documentos
Objetivo	Registrar información adicional para el programa de fiscalización, información necesaria para desarrollar los criterios de riesgo y/o modelos analíticos.
Actores	Jefe Unidad de Datos y Responsable del Programa
Precondiciones: el usuario debe estar autenticado en el perfil requerido y tener un programa de fiscalización vigente.	
Breve descripción: los perfiles autorizados suben o cargan información complementaria al programa de fiscalización en formatos electrónicos.	

Caso Uso 06	Cargar sujetos de fiscalización
Objetivo	Contar con el subconjunto de sostenedores y/o escuelas a los cuales aplica el programa de fiscalización.
Actores	Jefe Unidad de Datos y Responsable del Programa
Precondiciones: el usuario debe estar autenticado en el perfil requerido y tener un programa de fiscalización vigente.	
Breve descripción: los perfiles autorizados suben (individual o masivo) el listado de sostenedores y/o escuelas que conformarán el target u objetivo del programa de fiscalización, del cual se obtendrá la muestra a fiscalizar.	

Caso Uso 07	Ingresar los criterios de riesgo
Objetivo	Establecer los criterios de riesgo y propiedades que aplican al programa de fiscalización; junto con, cargar a cada uno los resultados.
Actores	Jefe Unidad de Datos, Responsable del Programa y Analista
Precondiciones: el usuario debe estar autenticado en el perfil requerido y tener un programa de fiscalización vigente.	
Breve descripción: los perfiles autorizados crean o editan cada criterio de riesgo consensuado para el programa de fiscalización; asignado un nombre, descripción ámbito que opera, fórmula de cálculo de la probabilidad e impacto y peso relativo. Además, el analista (fuera del sistema) calcula cada criterio por sujeto de fiscalización y tras ser visados (fuera del sistema) por el jefe unidad de datos, los carga en el sistema.	

Caso Uso 08	Calcular ranking
Objetivo	Obtener un listado ordenado de sostenedores y/o escuelas en función del grado de incumplimiento del programa.
Actores	Jefe Unidad de Datos, Responsable del Programa, Analista y Sistema
Precondiciones: el usuario debe estar autenticado en el perfil requerido, tener un programa de fiscalización vigente, los sujetos de fiscalización ingresados, los criterios de riesgo creados y la información de resultados de cada criterio ingresada.	
Breve descripción: los perfiles autorizados ejecutan el cálculo del ranking y revisan los resultados obtenidos, pudiendo aceptar los resultados o reprocesar.	

El siguiente esquema ilustra los casos de uso de la solución propuesta para el componente fichas de los programas de fiscalización.

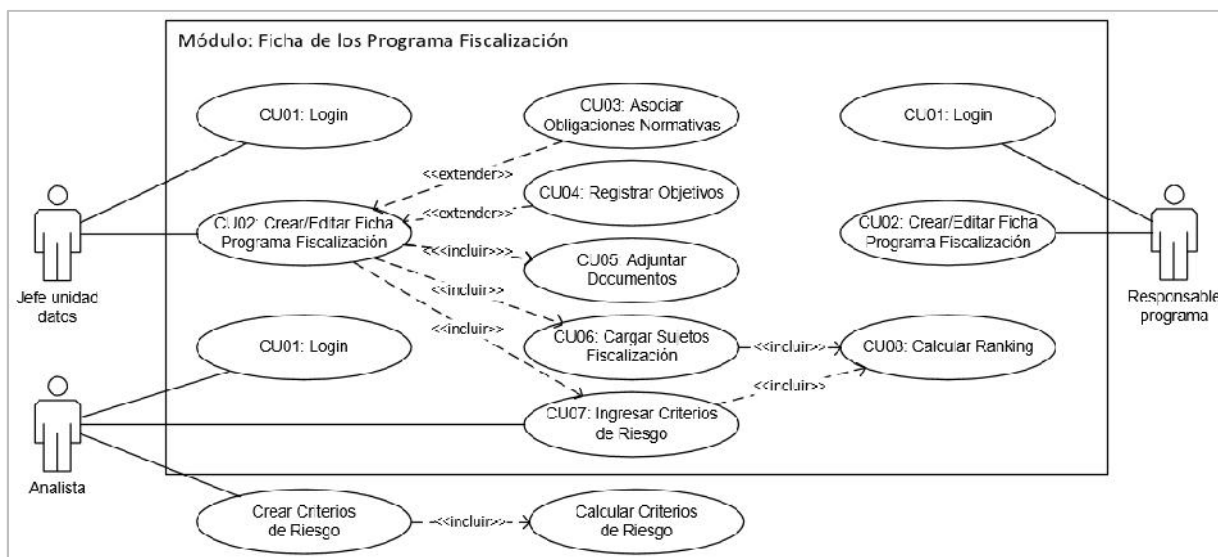


Figura N° 58: Casos de uso: módulo fichas de los programas fiscalización

### 6.3.2 Casos de Uso: Recomendaciones

Los siguientes casos de uso pertenecen a la etapa de preprocesamiento de textos.

Caso Uso 09	Cargar textos descripción gasto - cuenta
Objetivo	Identificar nuevos términos stopword, correctores ortográficos y n-grams para mantener la base de datos.
Actores	Analista negocio
Precondiciones: el usuario debe estar autenticado en el perfil requerido.	
Breve descripción: el analista de negocio carga en el sistema cada archivo de texto generado con por el proceso de extracción de datos que opera sobre el sistema de rendición de cuentas. El sistema devuelve un archivo de texto con aquellos términos no conocidos y donde el analista debe decidir si: es un término válido, stopword, correctores ortográficos o n-grams; para mantener la base de datos del sistema.	

Caso Uso 10	Crear reporte (.csv) términos
Objetivo	Facilitar la visualización de nuevos términos stopword, correctores ortográficos y n-grams.
Actores	Sistema
Precondiciones: contar con un archivo de texto (CU09) válido.	
Breve descripción: el sistema por cada fila del archivo de texto, crea un vector de términos tras tokenizarlo; luego cada término se consulta en la base de datos del sistema si es un stopword, corrector ortográfico y al unir dos o tres términos sucesivos, es un n-grams. El sistema por cada fila devuelve si el término está registrado.	

Caso Uso 11	Tokenizar texto
Objetivo	Crear un vector de término por cada campo de texto.
Actores	Sistema
Precondiciones: contar con un archivo de texto (CU09) válido.	
Breve descripción: el sistema por cada fila del archivo de texto y en particular, en el campo descripción del gasto se divide en un conjunto de términos ordenados por orden de aparición; donde, cada espacio en el relato del texto es el separador de términos.	

Caso Uso 12	Remover stopword
Objetivo	Reducir la cardinalidad del vector de término.
Actores	Sistema
Precondiciones: contar con un vector de términos válido.	
Breve descripción: el sistema busca en su tabla de datos de palabras de escaso contenido semántico cada término del vector, si el término es identificado es removido.	

Caso Uso 13	Corregir ortografía
Objetivo	Contar con vectores de términos correctamente escritos.
Actores	Sistema
Precondiciones: contar con un vector de términos válido.	
Breve descripción: el sistema busca en su tabla de datos de palabras “dice” (errores ortográficos) cada término del vector, si el término es identificado se modifica el término por el contenido “debe decir” de la tabla de datos.	



Caso Uso 14	Marcar n-grams
Objetivo	Unir grupos de términos que juntos poseen más significado que separados en el vector de términos.
Actores	Sistema
Precondiciones: contar con un vector de términos válido.	
Breve descripción: el sistema por cada término busca el siguiente y los une, luego revisa en su tabla de datos si el n-grams existe, de existir los trata como un solo. El proceso se repite ahora con tres términos sucesivos.	

Caso Uso 15	Descargar reporte (.csv) términos
Objetivo	Clasificar términos no reconocidos por las tablas de datos del sistema.
Actores	Analista negocio
Precondiciones: contar con un archivo de texto (.csv) a preprocesar.	
Breve descripción: el analista negocio revisa el archivo (.csv) preprocesado, tras ser tokenizado, removido los stopword, aplicado el corrector ortográfico e identificado los n-grams; todos aquellos términos no clasificados, el analista debe decidir si es un término válido o actualizar las respectivas tablas de datos.	

Caso Uso 16	Actualizar tabla stopword
Objetivo	Enriquecer la tabla de datos stopword
Actores	Analista negocio
Precondiciones:	

Breve descripción: el analista negocio utilizando el reporte provisto por el sistema o por iniciativa propia enriquece la tabla de datos stopword, sea por carga manual o carga masiva utilizando una plantilla (archivo.csv).

Caso Uso 17	Actualizar tabla stopword
Objetivo	Enriquecer la tabla corrector ortográfico
Actores	Analista negocio
Precondiciones:	
Breve descripción: el analista negocio utilizando el reporte provisto por el sistema o por iniciativa propia enriquece la tabla de datos corrector ortográfico, sea por carga manual o carga masiva utilizando una plantilla (archivo.csv).	

Caso Uso 18	Actualizar tabla n-grams
Objetivo	Enriquecer la tabla de datos n-grams
Actores	Analista negocio
Precondiciones:	
Breve descripción: el analista negocio utilizando el reporte provisto por el sistema o por iniciativa propia enriquece la tabla de datos n-grams, sea por carga manual o carga masiva utilizando una plantilla (archivo.csv).	

Caso Uso 19	Crear nuevo nodo término
Objetivo	Enriquecer la bases de datos y grafos del sistema
Actores	Analista negocio

Precondiciones:
Breve descripción: el analista negocio utilizando el reporte (.csv) o por iniciativa propia completa una plantilla (.csv) cada término es un nodo y enriquece las bases de datos y de grafos del sistema.

Caso Uso 20	Crear relaciones entre términos
Objetivo	Enriquecer la bases de datos y grafos del sistema con relaciones entre términos
Actores	Analista negocio
Precondiciones:	
Breve descripción: el analista negocio utilizando el reporte (.csv) o por iniciativa propia completa una plantilla (.csv) para relaciones de variante o forma verbal; donde, cada par de términos representa una relación del tipo señalado por la plantilla para enriquecer las bases de datos y de grafos del sistema.	

El siguiente esquema ilustra los casos de uso de la solución propuesta para la etapa de preprocesamiento de textos.

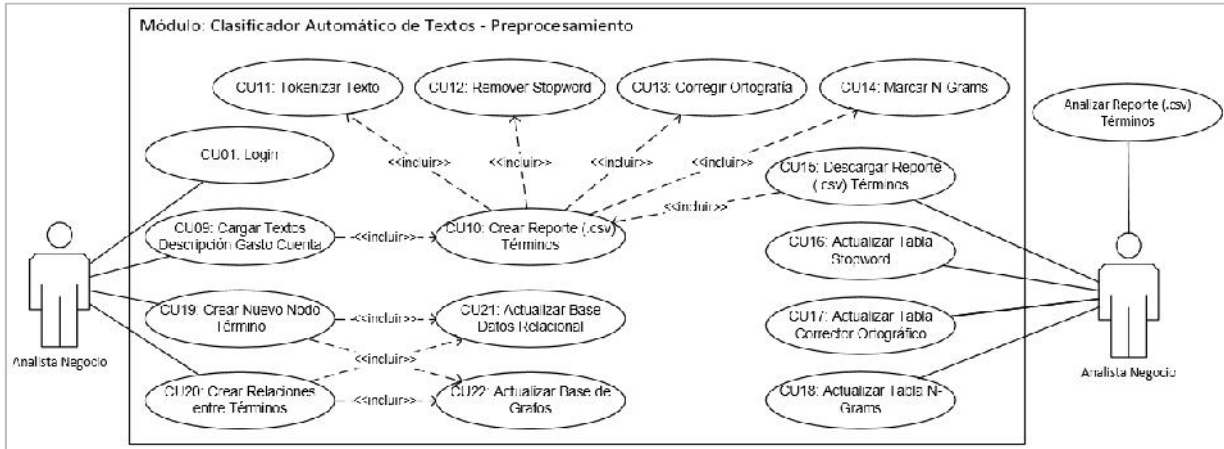


Figura N° 59: Casos de uso: módulo clasificador de textos - preprocesamiento

El siguiente esquema ilustra los casos de uso de la solución propuesta para la etapa de entrenamiento de la red de palabra para una subvención, cuenta imputación del gasto y tipo de recomendación.

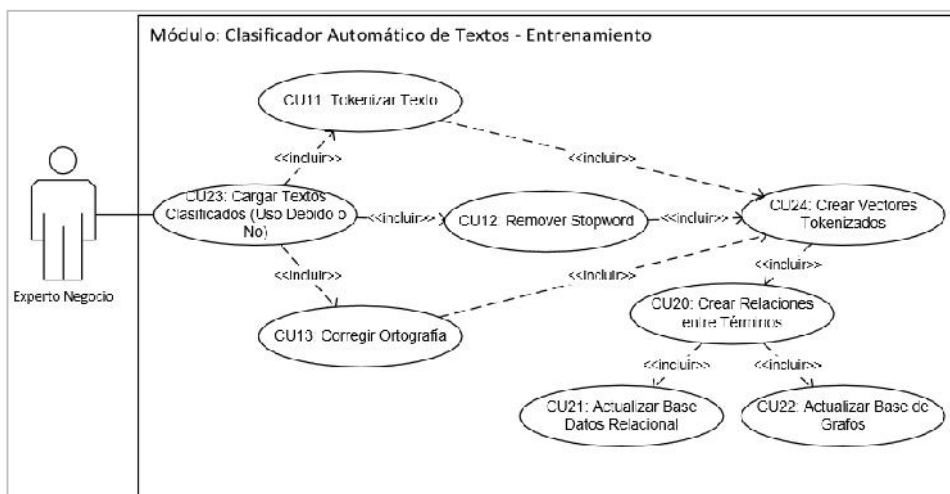


Figura N° 60: Casos de uso: módulo clasificador de textos – entrenamiento

El siguiente esquema ilustra los casos de uso de la solución propuesta para la etapa de clasificación para un texto según subvención, cuenta imputación del gasto y tipo de recomendación.

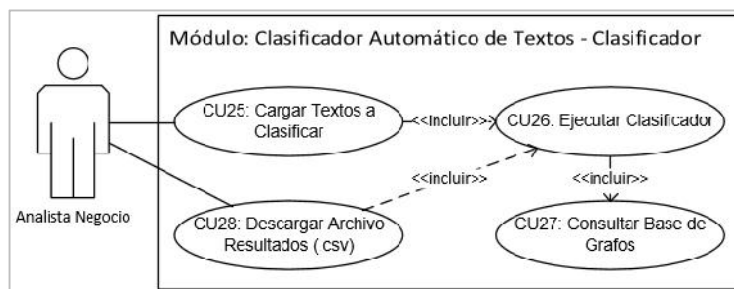


Figura N° 61: Casos de uso: módulo clasificador de textos – clasificador

### 6.3.3 Diagramas de Secuencia: Fichas de los Programas

Los siguientes esquemas ilustran los diagramas de secuencias de la solución propuesta.

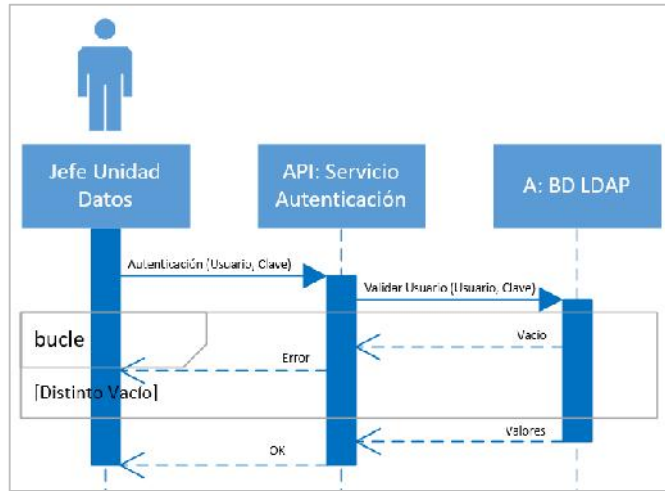


Figura N° 62: Diagrama secuencia CU01: login

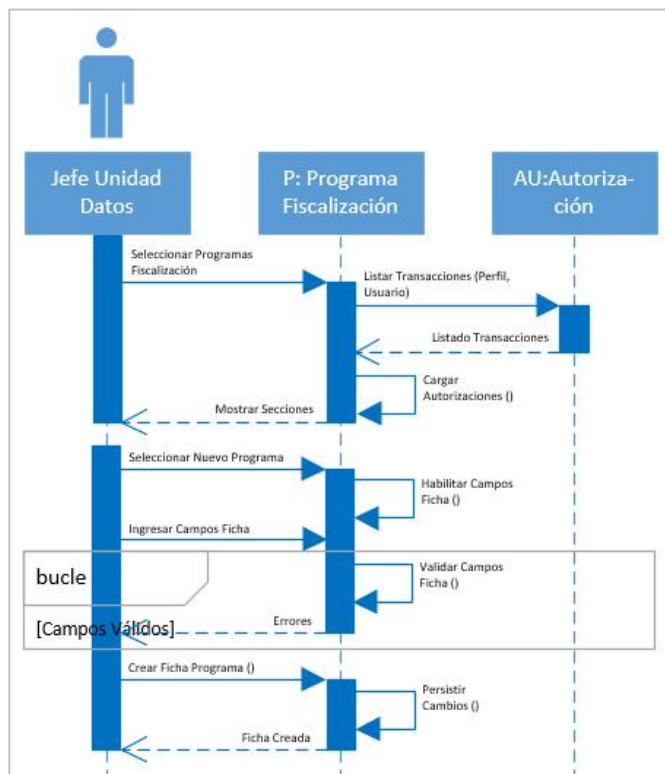


Figura N° 63: Diagrama secuencia CU02: crear/editar ficha programa fiscalización

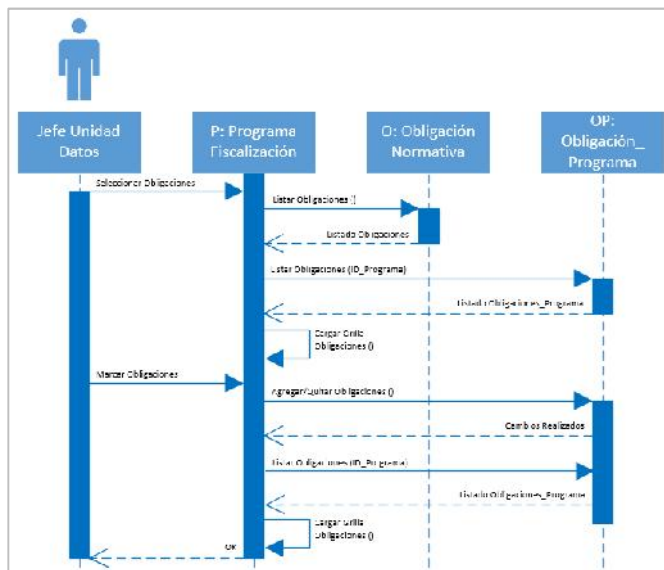


Figura N° 64: Diagrama secuencia CU03: asociar obligaciones normativas

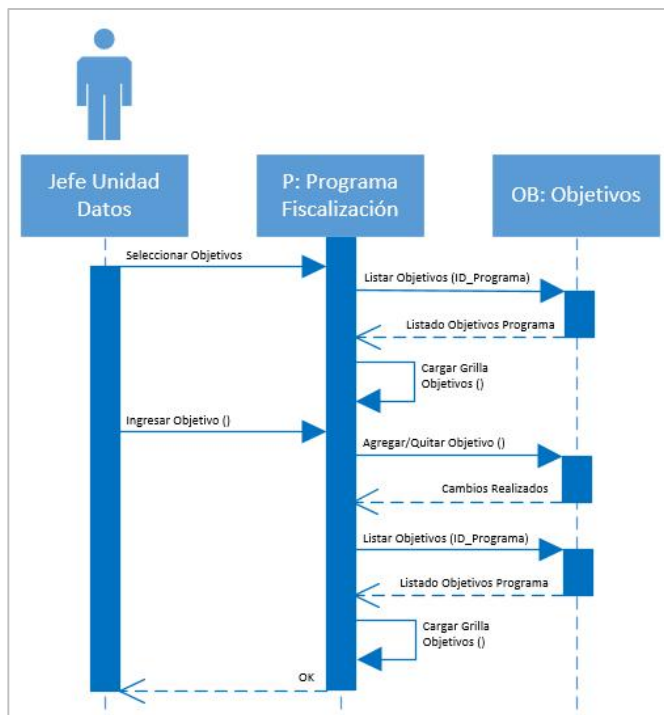


Figura N° 65: Diagrama secuencia CU04: registrar objetivos

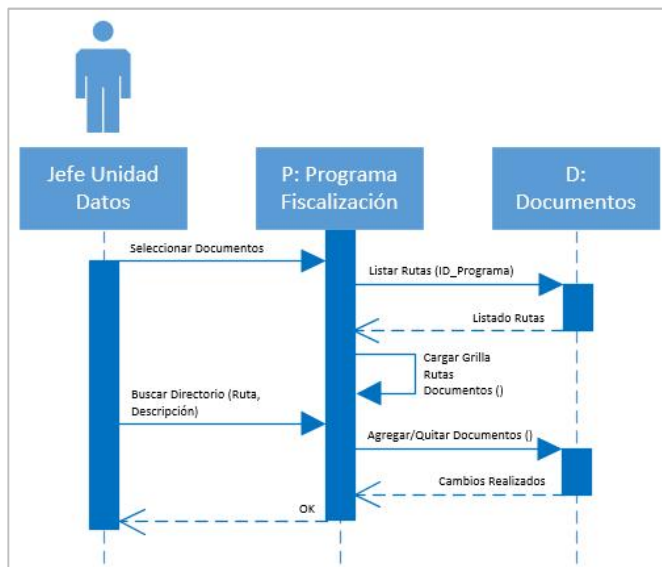


Figura N° 66: Diagrama secuencia CU05: adjuntar documentos

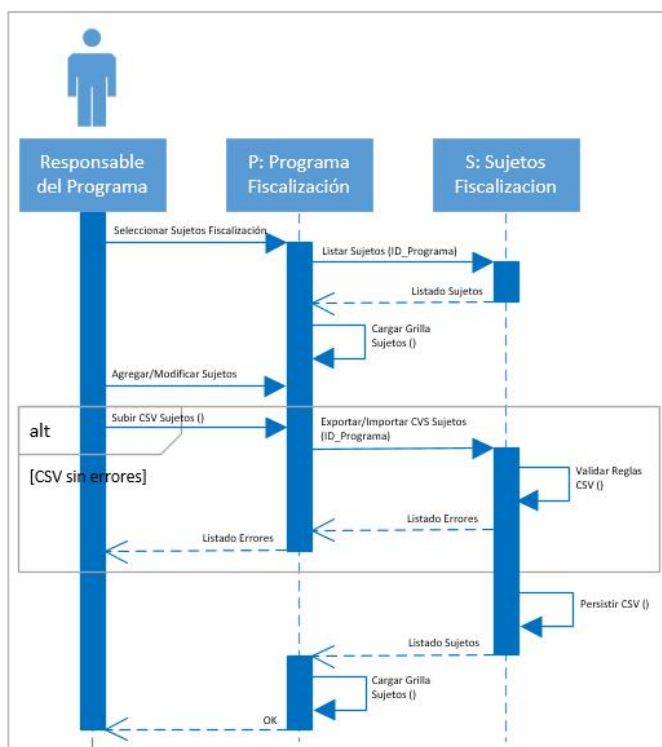


Figura N° 67: Diagrama secuencia CU06: cargar sujetos fiscalización

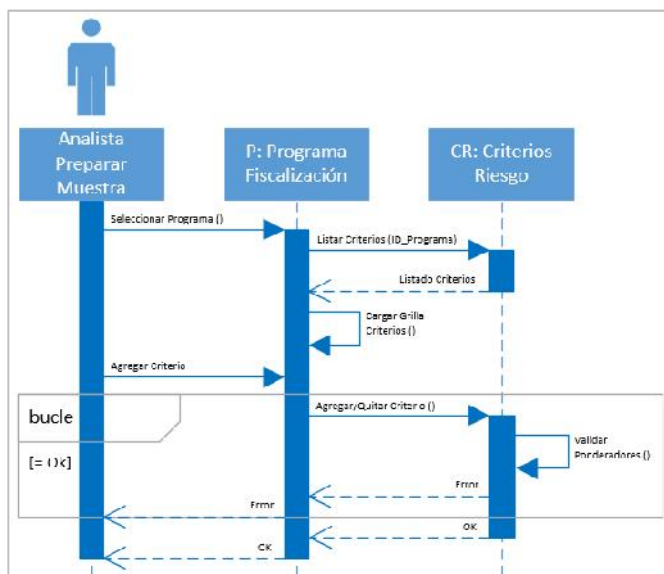


Figura N° 68: Diagrama secuencia CU07: ingresar criterios riesgo

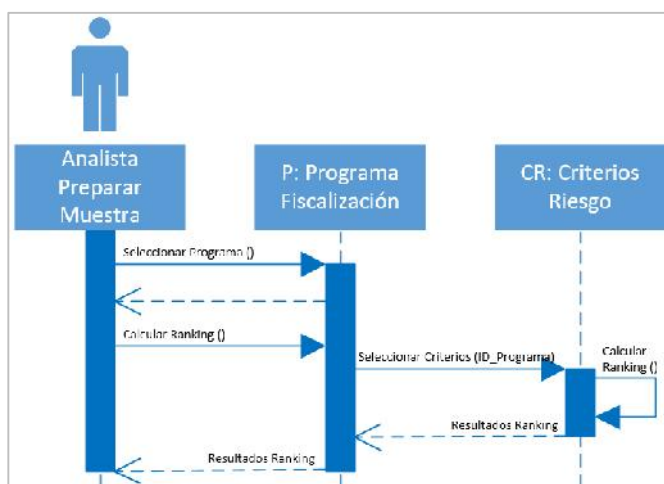


Figura N° 69: Diagrama secuencia CU08: calcular ranking



### 6.3.4 Diagramas de Secuencia: Recomendaciones

Los siguientes esquemas ilustran los diagramas de secuencias de la solución propuesta.

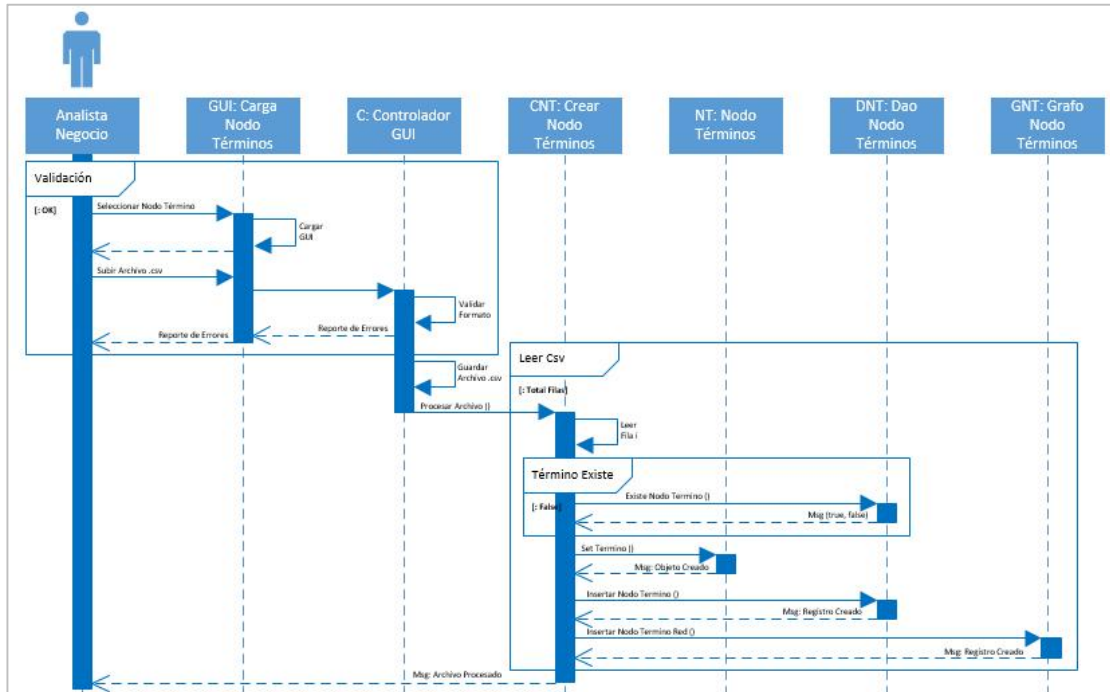


Figura N° 70: Diagrama secuencia CU19: crear nuevo nodo término

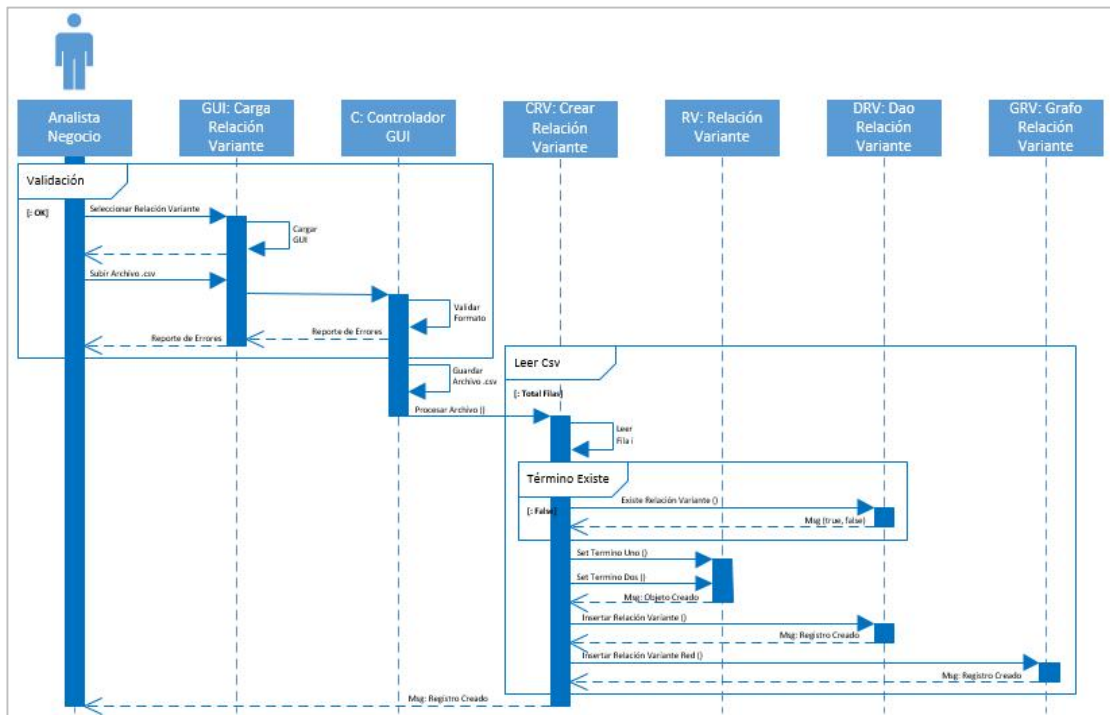


Figura N° 71: Diagrama secuencia CU20: crear relaciones entre términos

### 6.3.5 Diagramas de Clases: Ficha Programa

El siguiente esquema ilustra el diagrama de clases de la solución propuesta.

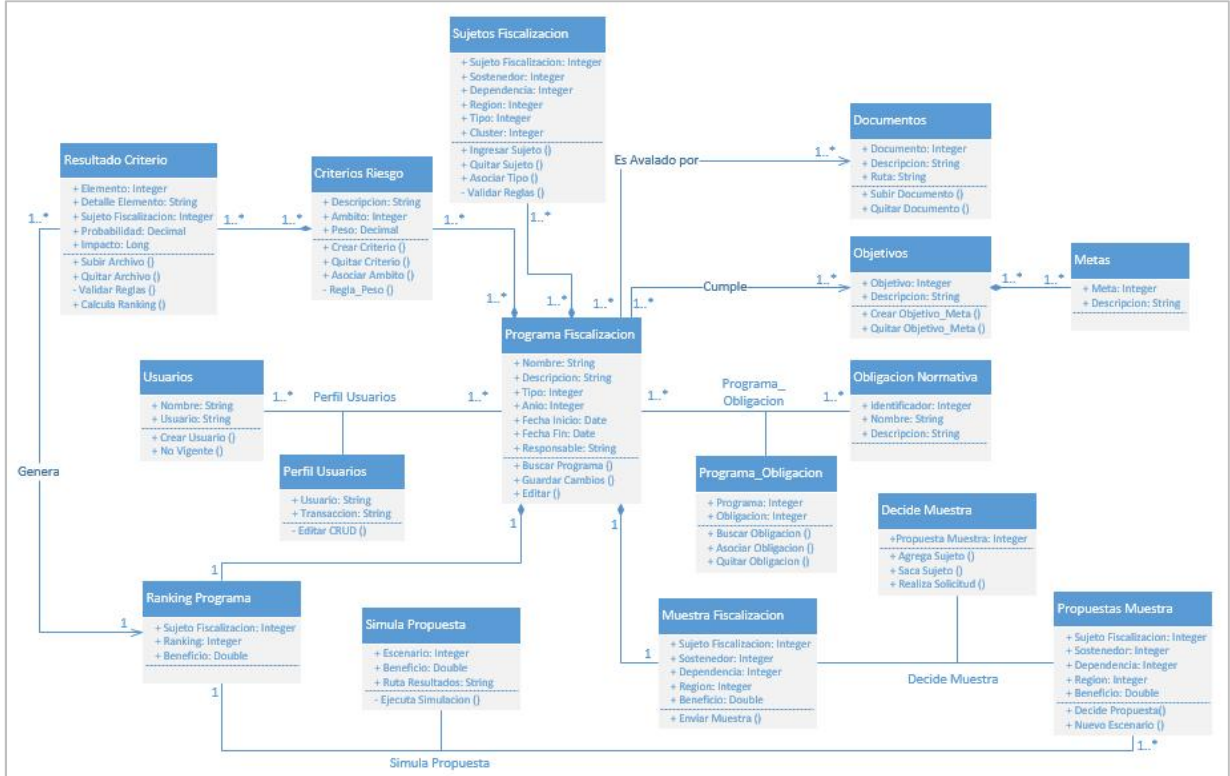


Figura N° 72: Diagrama de clases: ficha programas

### 6.3.6 Diagramas de Clases: Recomendaciones

El siguiente esquema ilustra el diagrama de clases de la solución propuesta.

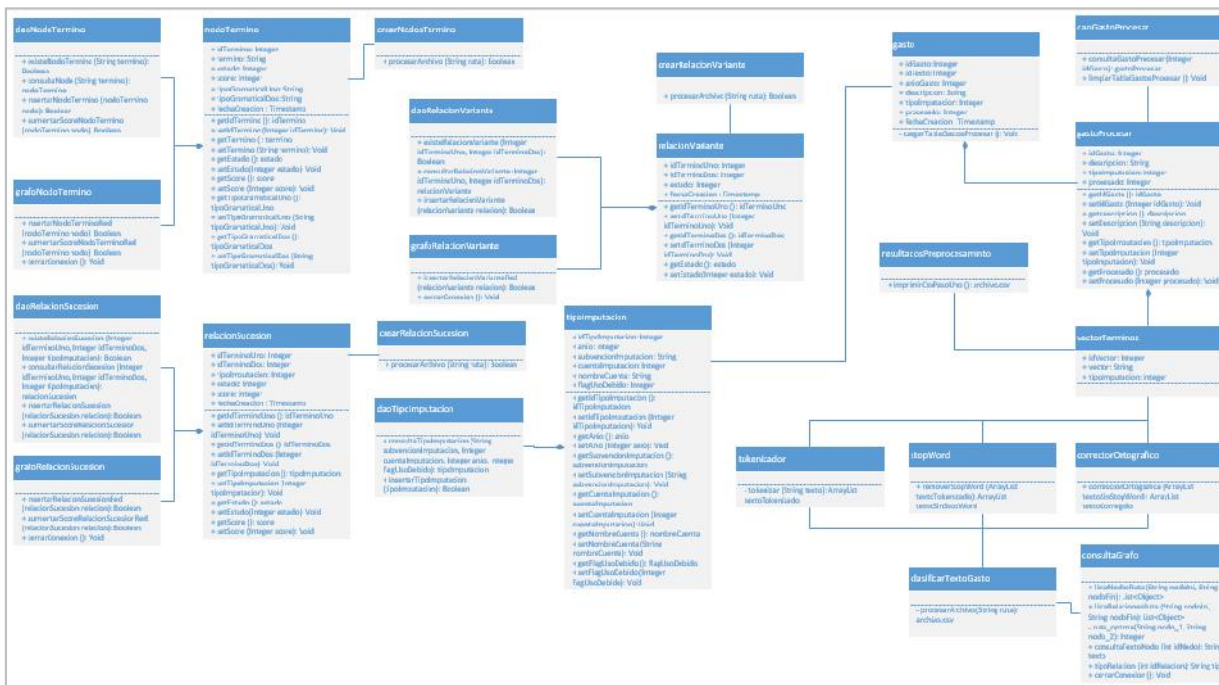


Figura N° 73: Diagrama de secuencias: recomendaciones

### 6.3.7 Prototipo Funcional Desarrollado: Ficha Programa

Los siguientes esquemas ilustran los mockup de la solución propuesta.



Figura N° 74: Autenticación

Programas de Fiscalización - Jefe Unidad de Datos      Búsqueda De Programas:       Buscar

Datos Generales    Obligaciones Normativas    Objetivos y Metas    Documentos

Descripción Programa:

Tipo de Programa: Normativa      #Año de Ejecución: 2016

Responsable del Programa: Seleccionar Usuario

Fecha de Inicio de Vigencia: dd-mm-aaaa      Fecha de Fin de Vigencia: dd-mm-aaaa

Guardar

Figura N° 75: Crear/editar programas de fiscalización: datos generales

Programas de Fiscalización - Jefe Unidad de Datos      Búsqueda De Programas:       Buscar

Datos Generales    Obligaciones Normativas    Objetivos y Metas    Documentos

Agregar Nueva Obligación

Obligaciones Normativas	Eliminar
<input type="text" value="Seleccione"/> <ul style="list-style-type: none"> <li>Seleccione</li> <li>Normativa 1</li> <li>Normativa 2</li> <li>Normativa 3</li> </ul>	<input type="button" value="-"/> <input type="button" value="Guardar"/>

Figura N° 76: Crear/editar programas de fiscalización: obligaciones normativas

Programas de Fiscalización - Jefe Unidad de Datos      Búsqueda De Programas:       Buscar

Datos Generales    Obligaciones Normativas    Objetivos y Metas    Documentos

Indicadores: Sin selección

Agregar Nuevo Objetivo

Objetivos del Programa	Editar	Eliminar
<input type="text" value="Objetivo"/>	<input type="button" value="✎"/>	<input type="button" value="-"/>

Agregar Nueva Meta

Metas del Programa	Editar	Eliminar
<input type="text" value="Meta"/>	<input type="button" value="✎"/>	<input type="button" value="-"/>

Guardar

Figura N° 77: Crear/editar programas de fiscalización: objetivos

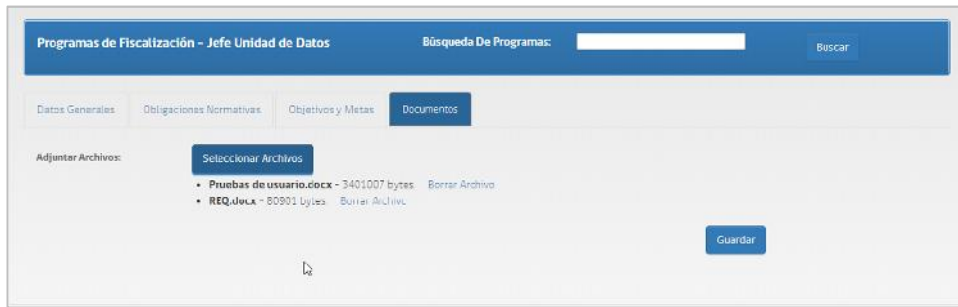


Figura N° 78: Crear/editar programas de fiscalización: cargar documentos

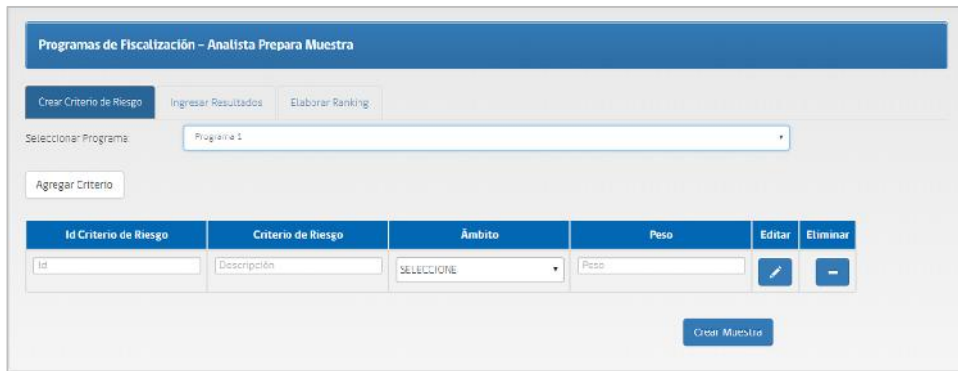


Figura N° 79: Ingresar criterios de riesgo: crear criterio

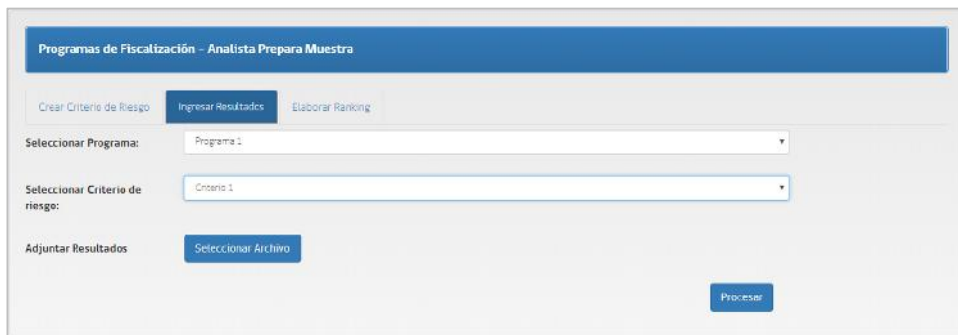


Figura N° 80: Ingresar criterios de riesgo: cargar resultados por criterio

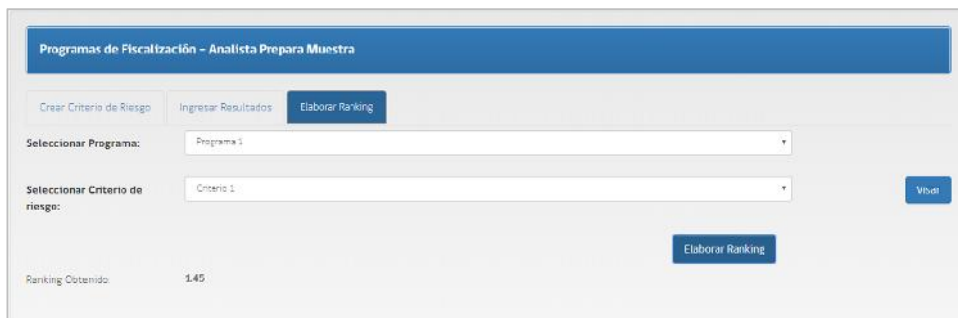


Figura N° 81: Calcular ranking

### 6.3.8 Prototipo Funcional Desarrollado: Recomendaciones

El siguiente componente (ver Figura N° 56) corresponde a la extracción de datos que se realiza periódicamente a la base de datos de rendición de cuentas.

La siguiente figura ilustra el ETL creado para estos fines.

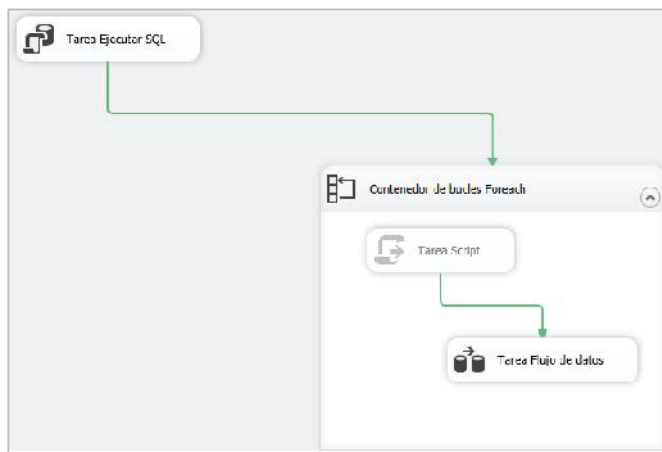


Figura N° 82: ETL extracción rendición de cuentas

La siguiente figura ilustra el detalle de la SQL de extracción de datos.

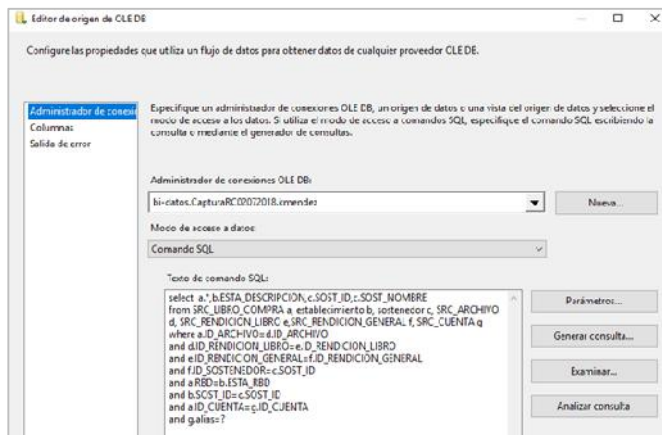


Figura N° 83: Contenido del origen DB utilizado

La siguiente figura ilustra el resultado del proceso, un archivo diario por cuenta de imputación en extensión .csv requerido por el clasificador automático de textos para preparar recomendaciones a cada fila del archivo.

ID	LIBRO	CUBRO	ID_SUBVENCIÓN	CUENTA	ID_EFO	DIGITALE	FECHA_DOCUMENTO	FECHA_PAGO	DESCRIPCION	PROVEEDOR	CANTIDAD	IMPORTE	CONTO	ARCHIVO	MONEDA	GAEDUC	FINANCIACION	CLAVE	DOCELEST	DESCRIBTO	SUBV	CONTRIBUCION	PROY	ID_CUENTA
2	102285	10256	5	533	27	223994	29-12-2017	00:00	COMBUSTIBLE	COMERCIAL	30000	15016	18000	22294	BO	COLEGIO	LOC	7630640	SANDRA ELIC COMBUSTIB	522			522	
3	303120	20002	5	538	20	235793	25-05-2017	00:00	COMBUSTIBLE	COMERCIAL	35115	34810	35110	18797	BO	ESCUELA	BAJ	7618918	SOCIEDAD RE COMBUSTIB	522			522	
4	364301	8313	5	533	20	968892	16-07-2017	00:00	COMBUSTIBLE	COMERCIAL	35800	48317	35000	99682	BO	ESCUELA	ING	6273800	FUNDACION COMBUSTIB	522			522	
5	1202111	16600	5	538	26	6884	24-10-2017	00:00	COMBUSTIBLE	COMERCIAL	14000	70366	18000	4896	BO	FACULTAD	ADM	7797900	COMPAÑIA COMBUSTIB	522			522	
6	1209767	18767	5	538	27	50442	01-01-2017	00:00	COMBUSTIBLE	COMERCIAL	3172	65300	3172	50442	BO	COLEGIO	MAJ	6039010	MIRAMAR COMBUSTIB	522			522	
7	1142284	20160	5	533	30	127749	06-07-2017	00:00	COMBUSTIBLE	COMERCIAL	36000	70211	36000	107749	BO	FACULTAD	ADM	6514014	CONTRATA COMBUSTIB	522			522	
8	1828118	71002	5	538	27	11386735	05-07-2017	00:00	COMBUSTIBLE	COMERCIAL	25000	111104	25000	11386735	BO	COMBUSTIBLE	ADM	6273800	FUNDACION COMBUSTIB	522			522	
9	1502793	7013	5	533	20	6756	10-11-2017	00:00	COMBUSTIBLE	COMERCIAL	72250	95346	72250	6756	BO	FACULTAD	ADM	6922401	ILUSTRACION COMBUSTIB	522			522	
10	3566895	7918	5	538	30	5093	05-05-2017	00:00	COMBUSTIBLE	COMERCIAL	30000	188416	30000	5093	BO	FACULTAD	ADM	7613524	SOCIEDAD RE COMBUSTIB	522			522	
11	3221279	22737	5	538	30	389299	22-09-2017	00:00	COMBUSTIBLE	COMERCIAL	25000	132572	25000	389299	BO	FACULTAD	ADM	6514014	CONTRATA COMBUSTIB	522			522	
12	3350111	20127	5	538	30	58156	30-05-2017	00:00	COMBUSTIBLE	COMERCIAL	205000	138703	205000	58156	BO	FACULTAD	ADM	6514014	CONTRATA COMBUSTIB	522			522	
13	4892969	3638	5	538	26	12000000	01-09-2017	00:00	COMBUSTIBLE	COMERCIAL	11000	48862	11000	12000000	BO	COMBUSTIBLE	ADM	6514014	CONTRATA COMBUSTIB	522			522	
14	3305140	6263	5	538	26	62864817	27-01-2017	00:00	COMBUSTIBLE	COMERCIAL	2400	333518	2400	62864817	BO	COMBUSTIBLE	ADM	6514014	CONTRATA COMBUSTIB	522			522	
15	3328939	11905	5	538	27	866070	08-09-2017	00:00	COMBUSTIBLE	COMERCIAL	30000	393933	30000	866070	BO	FACULTAD	ADM	7111260	INSTITUTO COMBUSTIB	522			522	
16	3186798	2492	5	538	20	22201091	01-12-2017	00:00	COMBUSTIBLE	COMERCIAL	700	174857	700	22201091	BO	FACULTAD	ADM	7111260	INSTITUTO COMBUSTIB	522			522	
17	3202137	11005	5	538	27	025607	04-12-2017	00:00	COMBUSTIBLE	COMERCIAL	36000	377215	36000	025607	BO	FACULTAD	ADM	7111260	INSTITUTO COMBUSTIB	522			522	
18	4808812	14348	5	538	26	1483153	31-10-2017	00:00	COMBUSTIBLE	COMERCIAL	2700	457223	2700	1483153	BO	COMBUSTIBLE	ADM	6507368	FUNDACION COMBUSTIB	522			522	
19	3541449	9560	5	538	27	156436	21-02-2017	00:00	COMBUSTIBLE	COMERCIAL	30000	338807	30000	156436	BO	COMBUSTIBLE	ADM	7183240	SOCIEDAD COMBUSTIB	522			522	
20	4706333	60	5	538	20	187684	16-12-2017	00:00	COMBUSTIBLE	COMERCIAL	5875	447975	5875	187684	BO	COMBUSTIBLE	ADM	6509401	MIP FUNDACION COMBUSTIB	522			522	
21	4929248	8070	5	538	30	394788	10-09-2017	00:00	COMBUSTIBLE	COMERCIAL	22333	406927	22333	394788	BO	FACULTAD	ADM	7613626	ESPECIACION COMBUSTIB	522			522	
22	5405567	11216	5	538	30	123339	20-09-2017	00:00	COMBUSTIBLE	COMERCIAL	30000	510329	30000	123339	BO	FACULTAD	ADM	6514014	CONTRATA COMBUSTIB	522			522	
23	581856	10842	5	538	27	764344	14-06-2017	00:00	COMBUSTIBLE	COMERCIAL	30000	465299	30000	764344	BO	FACULTAD	ADM	6509426	FUNDACION COMBUSTIB	522			522	
24	6260293	6252	5	538	20	347900	12-05-2017	00:00	COMBUSTIBLE	COMERCIAL	24700	541061	24700	347900	BO	COMBUSTIBLE	ADM	6514014	CONTRATA COMBUSTIB	522			522	
25	5731550	5803	5	538	30	317790	09-02-2017	00:00	COMBUSTIBLE	COMERCIAL	1300	528961	1300	317790	BO	COMBUSTIBLE	ADM	6509991	CONTRATA COMBUSTIB	522			522	
26	6350913	16477	5	538	20	344118	23-10-2017	00:00	COMBUSTIBLE	COMERCIAL	7000	584264	7000	344118	BO	COMBUSTIBLE	ADM	7893259	SOCIEDAD COMBUSTIB	522			522	
27	925454	9805	5	538	27	3306793	01-12-2017	00:00	COMBUSTIBLE	COMERCIAL	2435	39977	2435	3306793	BO	COMBUSTIBLE	ADM	6510259	FUNDACION COMBUSTIB	522			522	

Figura N° 84: Archivo de salida para clasificador automático de textos

Los siguientes mockups corresponden al módulo clasificador automático de textos.

Figura N° 85: Cargar archivo de texto a clasificar en una categoría

ID Documento	Texto	Resultado
1	Compra Combustible traslado alumnos	0,7
2	Compra de materiales de oficina	0,02

Figura N° 86: Resultados del clasificador textos para la categoría

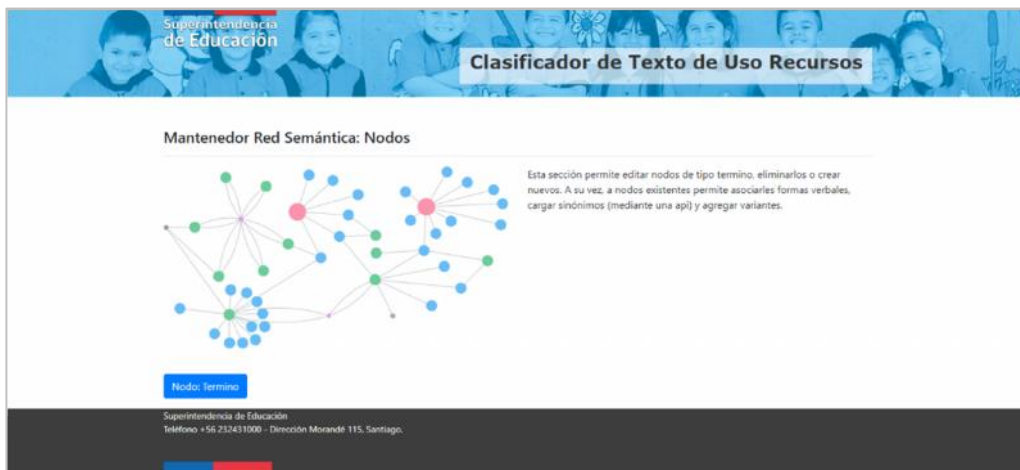


Figura N° 87: Mantenedor de nodos de la red semántica

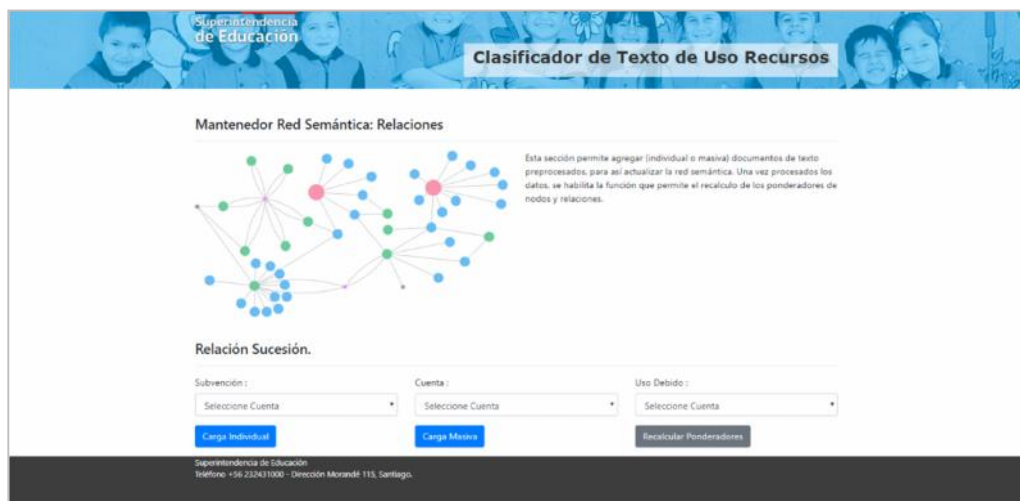


Figura N° 88: Mantenedor de relaciones de la red semántica

La siguiente figura ilustra la carga de recomendaciones en el sistema de rendición de cuentas.

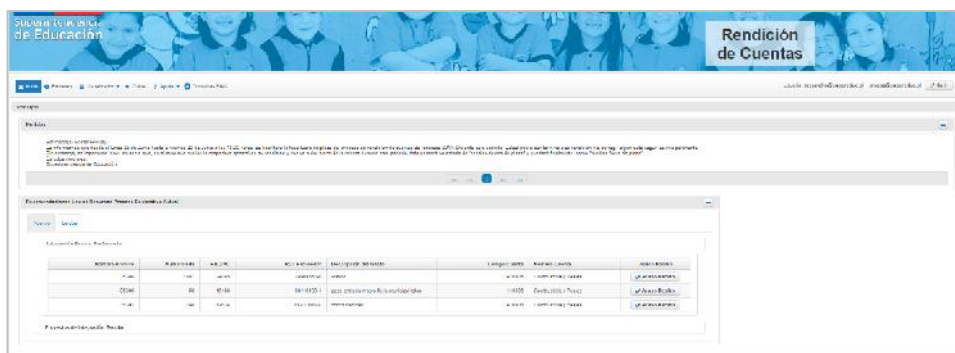


Figura N° 89: Publicación de recomendaciones: rendición de cuentas



La siguiente figura ilustra la carga de recomendaciones (nuevas materias) en el sistema de fiscalización (SIFE).

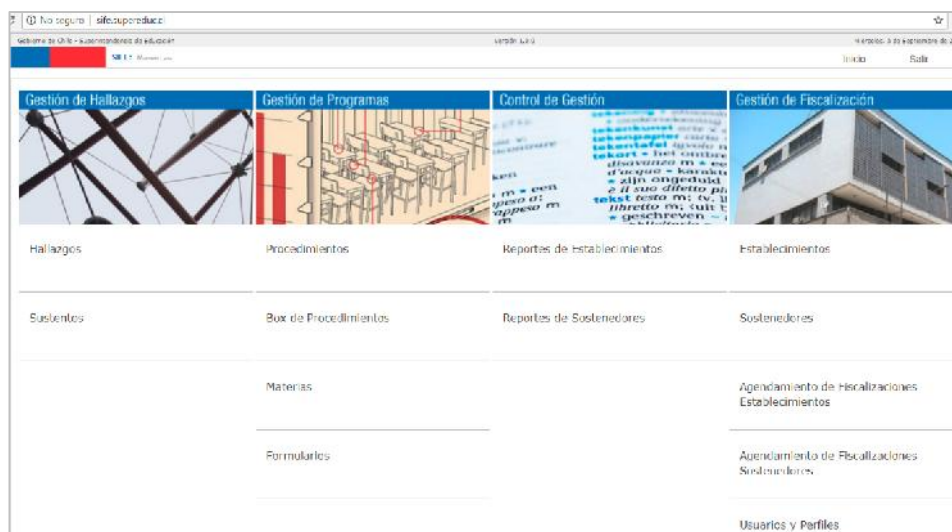


Figura N° 90: Carga de recomendaciones en SIFE

### 6.3.9 Datamart y Cubos OLAP

Los *datamart* son pequeños *datawarehouse* centrados en un tema o área de negocio en específico; el *datamart* de fiscalización, objeto del proyecto, actualiza sus datos en procesos batch diarios que consultan la base de datos transaccional del actual sistema de fiscalización, mediante la herramienta de Microsoft *integration services*.

A su vez, el *datamart* se consulta por medio de herramientas OLAP (*on line analytical processing*) compuestas de tablas de hechos (numéricos o medidas) y tablas de dimensiones, y sobre estas bases de datos se construyen los EIS (*executive information systems*, sistemas de información para directivos) y DSS (*decision support systems*, sistemas de ayuda a la toma de decisiones).

Las siguientes ilustraciones, responden a la implementación del *datamart* de fiscalización y sus cubos OLAP; estos últimos, responden a los indicadores de los flujos de información de control de avance de planificación para decisiones del proceso gestión de las fiscalizaciones, que apoyan a materializar el objetivo del proyecto de aumentar la actual capacidad de visitas de fiscalización para el programa uso de recursos SEP en un 2%.



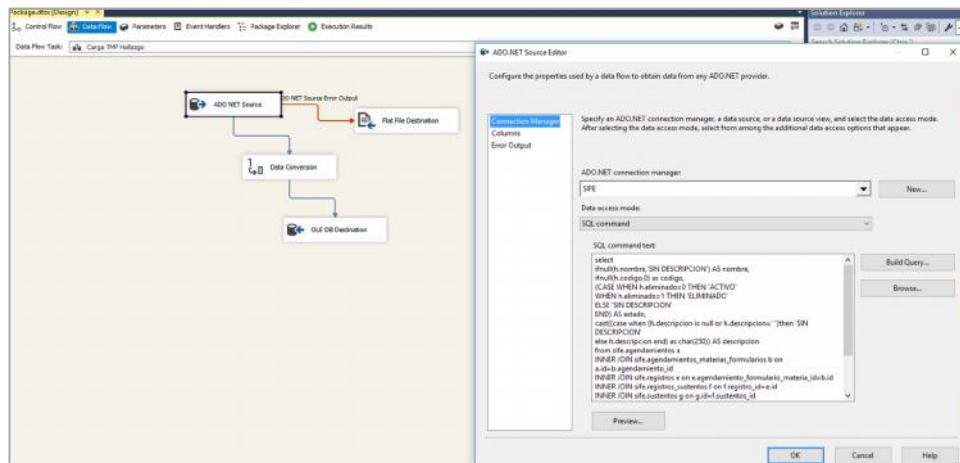


Figura N° 94: Detalle del operador: carga operador hallazgo

La siguiente figura presenta el resultado del cubo OLAP, objeto del proyecto, visto desde Excel; donde, el punto 1, señala la fecha del agendamento (gestión de las fiscalizaciones) y el punto 2 la fecha de registro del acta de fiscalización (ejecución de las fiscalizaciones).

Id	fecha_agendamento	fecha_abre_cispo_mcbil	fecha_cierra_cispo_mcbil	fecha_ingreso_web	nombre
1	2013-04-15 00:00:00	2013-04-15 16:39:11	2013-04-15 13:31:17	2013-04-15 13:32:05	Unidad de A a Z
2	2013-04-15 00:00:00				Unidad de Z a A
3	2013-04-15 00:00:00				Unidad por color
4	2013-04-15 00:00:00				Unidad filtro de "nombre"
5	2013-04-15 00:00:00				Filtros por región
6	2013-04-15 00:00:00				Filtros de texto
7	2013-04-15 00:00:00	2013-04-15 09:20:24	2013-04-16 07:21:06	2013-04-16 07:21:08	Busca:
8	2013-04-16 00:00:00				<input checked="" type="checkbox"/> PROG FISCALIZACION DE MIRA - MITA
9	2013-04-16 00:00:00				<input checked="" type="checkbox"/> PROG FISCALIZACION RECURSOS MIL 2016
10	2013-04-16 00:00:00				<input checked="" type="checkbox"/> PROG FISCALIZACION BARRANCA TRAFICAZO
11	2013-06-21 00:00:00	2013-07-01 08:55:59	2013-07-01 09:00:00	2013-07-01 09:16:20	<input checked="" type="checkbox"/> PROG RECURSOS SEP 2016
12	2013-06-21 00:00:00	2013-05-29 14:45:21	2013-06-26 15:14:07	2013-06-26 15:21:25	<input checked="" type="checkbox"/> PROG RECURSOS SEP 2015
13	2013-06-21 00:00:00	2013-04-29 13:28:18	2013-04-28 13:34:21	2013-04-28 13:32:28	<input checked="" type="checkbox"/> PROG REV SIMPOSIOS SEP RECURSOS 2015 2011
14	2013-06-21 00:00:00	2013-05-28 15:53:50	2013-06-28 16:51:59	2013-06-28 16:57:04	<input checked="" type="checkbox"/> PROG SEC BILIBIATO DECA Y REG ASISTENCIA 20 5
15	2013-06-21 00:00:00	2013-07-02 09:22:35	2013-07-02 09:24:10	2013-07-02 09:24:30	<input checked="" type="checkbox"/> PROG SEC BILIBIATO DECA Y REG ASISTENCIA 20 6
16	2013-06-21 00:00:00	2013-05-23 14:17:11	2013-06-28 15:15:15	2013-06-28 15:23:18	<input checked="" type="checkbox"/> PROG SEC BILIBIATO DECA Y REG ASISTENCIA 20 5
17	2013-06-21 00:00:00	2013-05-23 13:34:54	2013-06-28 13:37:15	2013-06-28 13:39:19	<input checked="" type="checkbox"/> PROG SEC BILIBIATO DECA Y REG ASISTENCIA 20 6
18	2013-06-21 00:00:00	2013-05-23 16:14:19	2013-06-28 16:10:18	2013-06-28 16:11:02	<input checked="" type="checkbox"/> PROG SEC BILIBIATO DECA Y REG ASISTENCIA 20 5
19	2013-06-21 00:00:00	2013-05-28 15:08:18	2013-06-28 15:18:47	2013-06-28 15:25:18	

Figura N° 95: Extracto base de cálculo indicador fecha agendamento v/s ejecución

De la figura anterior, por cada fila se obtiene la duración de la instrucción a fiscalizar (al restar ambas fechas) y la Figura N° 96 grafica el indicador: duración promedio de las visitas de fiscalización por programa; el cual, por cada día calendario registra la duración promedio de todas las fiscalizaciones ejecutadas ese día.

El indicador se entiende gestionado y optimizado, si los registros se ubican en mayoría en un rango o banda de días, tal como ilustra la Figura N° 96 entre los días 10 al 30 donde se efectúa control y las duraciones fluctúan entre 2 a 3 días; mientras que, en días posteriores donde no se realiza control, las duraciones fluctúan entre 4 a 10 días. Por lo

anterior, es factible aumentar la capacidad de visitas de fiscalización efectuando control a este indicador en a lo menos el porcentaje requerido como objetivo del proyecto.

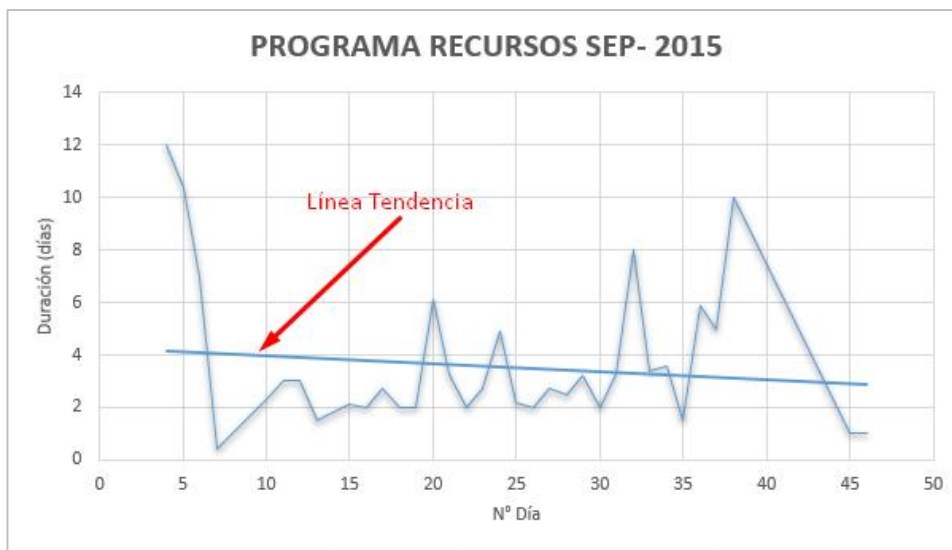


Figura N° 96: Indicador fecha agendamiento v/s ejecución, últimos 45 días

## **CAPÍTULO 7: EVALUACIÓN DEL PROYECTO**

### **7.1 PLAN PILOTO**

#### ***7.1.1 Definición del Plan Piloto***

Respecto a los modelos comportamientos de gestión financiera de sostenedores, éstos fueron piloteados con el comportamiento de Subvención Escolar Preferencial y datos históricos disponibles, tanto en la SUPEREDUC, MINEDUC y Agencia de la Calidad.

Respecto del clasificador automático de textos, fue piloteado con el programa uso de recursos SEP año 2016 (recursos año 2015) y la cuenta de gasto: combustible y peajes; esta fue aplicada a más de 40.000 documentos de textos, demostrando la validez del modelo. A su vez, los resultados de la clasificación, en particular los usos indebidos, se pilotearon para descubrir conocimiento de usos indebidos frecuentes.

#### ***7.1.2 Resultados Obtenidos***

Respecto a los modelos comportamientos de gestión financiera de sostenedores, se determinaron los clústeres de sostenedores con comportamientos riesgosos respecto de la gestión de recursos SEP; de forma tal, que cada sostenedor pertenece a un grupo y es posible tomar decisiones ad hoc a cada clúster. A su vez, tras aplicar métodos de clasificación se descubrieron patrones de comportamientos que serán incluidos en fiscalizaciones preventivas y acciones de difusión de información para instalar o cambiar prácticas en sostenedores.

Respecto de la clasificación automática de textos, tras entrenar el modelo con 3.000 documentos de textos, clasificaron más de 39.000 documentos de textos según su uso, demostrando su validez. A su vez, la información clasificada (uso indebido) fue piloteada para descubrir conocimiento de usos indebidos frecuentes con objeto de iniciar acciones de prevención y/o mitigación para instalar o cambiar prácticas en sostenedores.

## 7.2 DEFINICIÓN DE BENEFICIOS Y COSTOS

### 7.2.1 Beneficios

- a) Reintegro de Recursos al Sistema Escolar:

El Programa Uso Recursos SEP 2016, tuvo una capacidad de fiscalización restringida a 2.100 visitas a terreno y cada visita, en promedio, reintegró MM\$10,5. Además, se detectó MM\$ 76.131 en documentos riesgosos y se reintegró al sistema escolar MM\$ 22.251.

Si bien, el clasificador de textos detecta MM\$ 3.806 en documentos riesgosos, con el modelo de fiscalización actual, el monto del reintegro está sujeto a las visitas de fiscalización; por ello, el reintegro es MM\$ 441, que se explican al aumentar la capacidad de fiscalización para el programa en un 2% a partir del primer año.

- b) Ahorro Recursos por Buenas Prácticas:

Tal como fue descrito, al instalar buenas prácticas de uso debido para SEP, se reduce la actual tasa de 9,5% (monto documentos riesgosos/monto total SEP), llegando a un 7,5% en el segundo año y a un 6,5% en el tercero (base año 2016); esto adicional a la externalidad positiva para el sistema escolar, evita realizar 426 visitas a terreno en el segundo año (MM\$ 35) y 650 (MM\$ 54) para el tercero.

- c) Ahorro y Reintegro Recursos por Optimización:

Utilizar programación matemática y un software de apoyo para transformar el ranking de sostenedores y escuelas para el programa en la propuesta de muestra a fiscalizar, tal como se ha descrito reduce en un 70% los tiempos de proceso (MM\$ 0,7), debido a que es trabajado por un analista (sueldo 2 MM\$/mes) y le toma dos semanas. A su vez, el algoritmo permite aumentar en el primer año, un 0,5% del total reintegrado (MM\$ 112,5).

Tabla N° 22: Proyección de beneficios del proyecto

Año	a) Reintegro Recursos	b) Buenas Prácticas	c) Optimización	Total (MM\$)
1	MM\$ 441		MM\$ 0,7 y MM\$ 112,5	554
2		MM\$ 35	MM\$ 0,7 y MM\$ 88	124
3		MM\$ 54	MM\$ 0,7 y MM\$ 76	131

## 7.2.2 Costos

- Visita a Terreno de Fiscalización:

El Proyecto aumenta en un 2% la capacidad de visitas de fiscalización para el programa uso de recursos SEP (42 visitas) utilizando la capacidad ociosa y solo incurre en el costo de movilización MM\$ 0,22 en el primer año.

En base a la tabla N° 19, el costo de una visita de fiscalización<sup>27</sup> es \$ 84.170.

Tabla N° 23: Costo visita a terreno de fiscalización

Fiscalizador		Movilización	
Item	Cantidad	Item	Cantidad
Sueldo \$ / Mes	1.800.000	Bencina Anual \$	75.000.000
Hora Trabajo / Mes	190	Visita Anual #	21.150
Sueldo \$ / Hora Trabajo	9.500	Mantenimiento Anual \$	35.800.000
Duración Hora / Visita	8,3	Movilización \$ Anual	110.800.000
<b>Costo \$ / Visita</b>	<b>78.850</b>	<b>Movilización \$ / Visita</b>	<b>5.239</b>
		<b>Visita \$</b>	<b>84.089</b>

- Diseño e Implementación Algoritmo Programación Matemática:

Trabajo realizado por un analista (sueldo 2 MM\$/mes) y le toma dos semanas, totalizando un costo de MM\$ 0,7.

- Diseño e Implementación Clasificador de Textos Uso de Recursos:

Trabajo realizado por un experto en data mining (sueldo 3 MM\$/mes), analista de datos (sueldo 2 MM\$/mes), experto en uso de recursos (sueldo 2,5 MM\$/mes) y desarrollador Java (sueldo 1,8 MM\$/mes); y les toma tres meses, totalizando un costo de MM\$ 27,9.

---

<sup>27</sup> Los datos de sueldos de fiscalizadores y gastos de movilización se obtienen de: <http://transparencia.supereduc.cl/> y <http://www.dipres.gob.cl/574/w3-multipropertyvalues-14437-2940.html>

- Preparación Términos de Referencia<sup>28</sup>:

El Proyecto Data Mart (orden de compra 709603-538-CM17) tomó una semana de un analista tecnológico (sueldo 1,8 MM\$/mes) y de un analista de compras (sueldo 1 MM\$/mes), totalizando un costo de MM\$ 0,7.

El Proyecto Diseño, Construcción e Implementación del Nuevo Sistema de Fiscalización (orden de compra 709603-58-SE18) tomó un mes de un analista tecnológico (sueldo 1,8 MM\$/mes) y una semana de un analista de compras (sueldo 1 MM\$/mes), totalizando un costo de MM\$ 2,05.

- Profesionales SUPEREDUC:

La siguiente tabla resumen los tiempos utilizados de los perfiles de la SUPEREDUC requeridos por el Proyecto, totalizando un costo de MM\$ 18,9.

Tabla N° 24: Estimación del costo de perfiles SUPEREDUC

Perfil	Mes 1					Mes 2					Mes 3					Mes 4					Mes 5					Total Horas	Costo (\$)					
	S1	S2	S3	S4	S5	S1	S2	S3	S4	S5	S1	S2	S3	S4	S5	S1	S2	S3	S4	S5	S1	S2	S3	S4	S5							
Superintendente	2					2					2					2					2					2					12	378.947
PMO	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	25	328.947
Jefe Fiscalización	2					2					2					2					2					2					12	252.632
Jefe Transparencia Financiera	2					2					2					2					2					2					12	221.053
Coordinador TIC	2					2					2					2					2					2					12	221.053
Usuario Líder	19	19	19	19	19	19	19	19	19	19	19	19	19	19	19	19	19	19	19	19	19	19	19	19	19	19	19	19	19	19	475	5.000.000
Jefe Proyecto	38	38	38	38	38	38	38	38	38	38	38	38	38	38	38	38	38	38	38	38	38	38	38	38	38	38	38	38	38	38	950	12.500.000
																										<b>Total (\$)</b>	<b>18.902.632</b>					

Nota 1: 190 horas/mes

Nota 2: Sueldos en [http://transparencia.supereduc.cl/iris\\_gplus/Transparencia/indexa.jsp](http://transparencia.supereduc.cl/iris_gplus/Transparencia/indexa.jsp)

- Mantenimiento Anual:

Se estima una mantención anual equivalente a un 15% de la inversión inicial, a partir del segundo año totalizando un costo de MM\$ 6, para años siguientes aumenta a tasa 10%.

<sup>28</sup> Revisar <http://www.mercadopublico.cl/Portal/Modules/Site/Busquedas/BuscadorAvanzado.aspx?q=2>



### 7.2.3 Inversiones

- Desarrollo de Software:

La iniciativa Data Mart tuvo un coste de MM\$ 44, pero solo un 15% de este monto es imputado al Proyecto (MM\$ 6,6).

La iniciativa Diseño, Construcción e Implementación del Nuevo Sistema de Fiscalización tuvo un coste de MM\$ 220, pero un 5% del monto es imputado al Proyecto (MM\$ 11).

- Licencias de Software:

El Proyecto utiliza las licencias: ibm ilog cplex optimization studio (coste anual de MM\$ 6,6) e ibm spss modeler profesional (coste anual de MM\$ 11,7), totalizando MM\$ 18,3.

- Hardware:

Se considera adquirir un servidor de datos con un valor aproximado de MM \$ 2,5 y un laptop para procesar datos de un valor aproximado de MM\$ 1,4; totalizando MM\$ 3,9.

## 7.3 FLUJO DE CAJA

### 7.3.1 Depreciaciones

La siguiente tabla resume las depreciaciones de los componentes del proyecto,

Tabla N° 25: Depreciaciones de los activos del proyecto

Item	Valor Inicial \$	Años Depreciación	Depreciación Anual
Servidor	2.500.000	6	416.667
Laptop	1.400.000	6	233.333
Software	17.600.000	5	3.520.000

### 7.3.2 Tasa de Descuento Privada

La información de la Tasa de Descuento del Proyecto, fue obtenida del trabajo de Araya (2012), el cual hace referencia a una institución fiscalizadora del Estado y señala:

- Tasa de Descuento Social<sup>29</sup>: 8%.
- Tasa de Descuento Privada: 26%.  $T \quad K_e = R_f + B \quad (R_m + R_f)$

La siguiente Tabla N° 26 ilustra el Flujo de Caja Privado para el Proyecto, de la cual se desprende un VAN (26%) = MM\$ 496.

Tabla N°26: Flujo de caja del proyecto

	Año 0	Año 1	Año 2	Año 3	Año 4	Año 5	Año 6
<b>Ingresos</b>		554.000.000	124.000.000	131.000.000			
+ Ingresos		554.000.000	124.000.000	131.000.000			
+ Intereses por Depósitos							
+ Otros Ingresos							
+/- Ganancias/Pérdidas de Capital							
<b>costos</b>		54.640.000	10.170.000	10.770.000			
- Costos Fijos							
- Costos Variables							
- Gastos Generales		50.470.000					
- Gastos Mantención			6.000.000	6.600.000			
- Pago de Intereses por Créditos							
- Depreciaciones Legales(*)		4.170.000	4.170.000	4.170.000	4.170.000	4.170.000	650.000
- Pérdidas del Ejercicio Anterior							
= Utilidad Antes de Impuestos		499.360.000	113.830.000	120.230.000	- 4.170.000	- 4.170.000	- 650.000
- Impuesto de Primera Categoría (25%)							
= Utilidad Después de Impuestos		499.360.000	113.830.000	120.230.000	- 4.170.000	- 4.170.000	- 650.000
+ Depreciaciones Legales		4.170.000	4.170.000	4.170.000	4.170.000	4.170.000	650.000
+ Pérdidas del Ejercicio Anterior							
+/- Ganancias/Pérdidas de Capital							
= Flujo de Caja Operacional		503.530.000	118.000.000	124.400.000	-	-	-
- Inversión Fija	39.800.000						
+ Valor Residual de los Activos							
- Capital de Trabajo							
+ Recuperación del Capital de Trabajo							
+ Préstamos							
- Amortizaciones							
= Flujo de Capitales	- 39.800.000	503.530.000	118.000.000	124.400.000	-	-	-
= Flujo de Caja Privado	- 39.800.000	503.530.000	118.000.000	124.400.000	-	-	-
<b>VAN (26%)</b>		496.341.319					

### 7.4 ANÁLISIS DE SENSIBILIDAD

El factor crítico del Proyecto, desde el punto de vista evaluación privada, radica en el beneficio "Reintegro de Recursos"; el cual, bajo la actual modalidad de ejecución del servicio a fiscalización, está restringido a la capacidad de fiscalización y a la maximización que genera el algoritmo de optimización.

<sup>29</sup> A la Fecha la tasa de descuento social no ha variado respecto de lo señalado por Araya (2012).

- Escenario: sin aumento de capacidad de fiscalización, lo demás constante

La siguiente Tabla N° 27 ilustra el Flujo de Caja Privado para el Proyecto, de la cual se desprende un VAN (26%) = MM\$ 146.

Tabla N° 27: Flujo de caja escenario sin visitas de fiscalización

	Año 0	Año 1	Año 2	Año 3	Año 4	Año 5	Año 6
<b>Ingresos</b>		113.000.000	124.000.000	131.000.000			
+ Ingresos		113.000.000	124.000.000	131.000.000			
+ Intereses por Depósitos							
+ Otros Ingresos							
+/- Ganancias/Pérdidas de Capital							
<b>costos</b>		54.640.000	10.170.000	10.770.000			
- Costos Fijos							
- Costos Variables							
- Gastos Generales		50.470.000					
- Gastos Mantención			6.000.000	6.600.000			
- Pago de Intereses por Créditos							
- Depreciaciones Legales(*)		4.170.000	4.170.000	4.170.000	4.170.000	4.170.000	650.000
- Pérdidas del Ejercicio Anterior							
<b>= Utilidad Antes de Impuestos</b>		58.360.000	113.830.000	120.230.000	- 4.170.000	- 4.170.000	- 650.000
- Impuesto de Primera Categoría (25%)							
<b>= Utilidad Después de Impuestos</b>		58.360.000	113.830.000	120.230.000	- 4.170.000	- 4.170.000	- 650.000
+ Depreciaciones Legales		4.170.000	4.170.000	4.170.000	4.170.000	4.170.000	650.000
+ Pérdidas del Ejercicio Anterior							
+/- Ganancias/Pérdidas de Capital							
<b>= Flujo de Caja Operacional</b>		62.530.000	118.000.000	124.400.000	-	-	-
- Inversión Fija	39.800.000						
+ Valor Residual de los Activos							
- Capital de Trabajo							
+ Recuperación del Capital de Trabajo							
+ Préstamos							
- Amortizaciones							
<b>= Flujo de Capitales</b>	- 39.800.000	62.530.000	118.000.000	124.400.000	-	-	-
<b>= Flujo de Caja Privado</b>	- 39.800.000	62.530.000	118.000.000	124.400.000	-	-	-
<b>VAN (26%)</b>	146.341.319						

- Escenario: sin algoritmos de optimización, lo demás constante

La siguiente Tabla N° 28 ilustra el Flujo de Caja Privado para el Proyecto, de la cual se desprende un VAN (26%) = MM\$ 263.

Tabla N° 28: Flujo de caja escenario sin algoritmo de optimización

	Año 0	Año 1	Año 2	Año 3	Año 4	Año 5	Año 6
<b>Ingresos</b>		441.000.000	-	-			
+ Ingresos		441.000.000					
+ Intereses por Depósitos							
+ Otros Ingresos							
+/- Ganancias/Pérdidas de Capital							
<b>costos</b>		54.640.000	10.170.000	10.770.000			
- Costos Fijos							
- Costos Variables							
- Gastos Generales		50.470.000					
- Gastos Mantención			6.000.000	6.600.000			
- Pago de Intereses por Créditos							
- Depreciaciones Legales(*)		4.170.000	4.170.000	4.170.000	4.170.000	4.170.000	650.000
- Pérdidas del Ejercicio Anterior							
<b>= Utilidad Antes de Impuestos</b>		386.360.000	- 10.170.000	- 10.770.000	- 4.170.000	- 4.170.000	- 650.000
- Impuesto de Primera Categoría (25%)							
<b>= Utilidad Después de Impuestos</b>		386.360.000	- 10.170.000	- 10.770.000	- 4.170.000	- 4.170.000	- 650.000
+ Depreciaciones Legales		4.170.000	4.170.000	4.170.000	4.170.000	4.170.000	650.000
+ Pérdidas del Ejercicio Anterior							
+/- Ganancias/Pérdidas de Capital							
<b>= Flujo de Caja Operacional</b>		390.530.000	- 6.000.000	- 6.600.000	-	-	-
- Inversión Fija	39.800.000						
+ Valor Residual de los Activos							
- Capital de Trabajo							
+ Recuperación del Capital de Trabajo							
+ Préstamos							
- Amortizaciones							
<b>= Flujo de Capitales</b>	- 39.800.000	390.530.000	- 6.000.000	- 6.600.000	-	-	-
<b>= Flujo de Caja Privado</b>	- 39.800.000	390.530.000	- 6.000.000	- 6.600.000	-	-	-
<b>VAN (26%)</b>	263.065.775						

Flujo de Caja Privado: Escenario Pesimista (ingresos del primer se mantienen)

Tabla N° 29: Flujo de Caja Pesimista

	Año 0	Año 1	Año 2	Año 3	Año 4	Año 5	Año 6
Visitas con Resultados/año		2.996	2.996	2.996	2.996	2.996	2.996
Ingreso		84.170	84.170	84.170	84.170	84.170	84.170
+ Ingresos		252.173.320	252.173.320	252.173.320	252.173.320	252.173.320	252.173.320
+ Intereses por Depósitos							
+ Otros Ingresos							
+/- Ganancias/Pérdidas de Capital							
costo							
- Costos Fijos		95.500.000	98.365.000	101.315.950	104.355.429	107.486.091	110.710.674
- Costos Variables		-	-	-	-	-	-
- Gastos Generales		-	-	-	-	-	-
- Gastos Mantenición		15.000.000	16.500.000	18.150.000	19.965.000	21.961.500	24.157.650
- Pago de Intereses por Créditos							
- Depreciaciones Legales(*)		21.208.000	21.208.000	21.208.000	21.208.000	21.208.000	1.300.000
- Pérdidas del Ejercicio Anterior							
<b>= Utilidad Antes de Impuestos</b>		<b>120.465.320</b>	<b>116.100.320</b>	<b>111.499.370</b>	<b>106.644.892</b>	<b>101.517.729</b>	<b>116.004.996</b>
- Impuesto de Primera Categoría (25%)							
<b>= Utilidad Después de Impuestos</b>		<b>120.465.320</b>	<b>116.100.320</b>	<b>111.499.370</b>	<b>106.644.892</b>	<b>101.517.729</b>	<b>116.004.996</b>
+ Depreciaciones Legales		21.208.000	21.208.000	21.208.000	21.208.000	21.208.000	1.300.000
+ Pérdidas del Ejercicio Anterior							
+/- Ganancias/Pérdidas de Capital							
<b>= Flujo de Caja Operacional</b>		<b>141.673.320</b>	<b>137.308.320</b>	<b>132.707.370</b>	<b>127.852.892</b>	<b>102.817.729</b>	<b>116.004.996</b>
- Inversión Fija	155.631.579						
+ Valor Residual de los Activos							
- Capital de Trabajo							
+ Recuperación del Capital de Trabajo							
+ Préstamos							
- Amortizaciones							
<b>= Flujo de Capitales</b>	- 155.631.579	<b>141.673.320</b>	<b>137.308.320</b>	<b>132.707.370</b>	<b>127.852.892</b>	<b>102.817.729</b>	<b>116.004.996</b>
<b>= Flujo de Caja Privado</b>	- 155.631.579	<b>141.673.320</b>	<b>137.308.320</b>	<b>132.707.370</b>	<b>127.852.892</b>	<b>102.817.729</b>	<b>116.004.996</b>
<b>VAN (26%)</b>		<b>221.728.254</b>					

En conclusión, desde el punto de vista de evaluación privada, es recomendable ejecutar el proyecto. El flujo de caja original y sus escenarios resultan en un VAN  $\geq 0$ .

## **CAPÍTULO 8: CONCLUSIONES**

### **8.1 SOBRE LA METODOLOGÍA DE INGENIERÍA DE NEGOCIOS**

La metodología de ingeniería de negocios, mediante su enfoque *top-down*, permite tener una visión sistémica de la SUPEREDUC que facilita entender su funcionamiento y detectar problemáticas en áreas críticas. Junto con, proponer soluciones que se encuentren alineadas a estrategia organizacional, modelo de negocio, capacidades requeridas, procesos y tecnologías de apoyo.

Instanciando la metodología, se identificó que la SUPEREDUC ante el objetivo de *mejorar el uso de los recursos asignados para fines educativos*, sus procesos de fiscalización se activan en fecha posterior a la declaración anual de gastos, y durante el año no se tomaban medidas preventivas y/o de mitigación de riesgos derivados de usos indebidos de recursos. Con esta forma de operar, la SUPEREDUC detecta miles de millones de pesos con usos indebidos. Lo anterior, no se encuentra alineado al propósito de la SUPEREDUC de ser garante de derechos educativos, y en especial a cautelar el debido uso de los fondos de la subvención escolar preferencial (SEP) -recursos que por Ley deben destinarse a apoyar a los estudiantes más vulnerables del país y a garantizar que los proyectos educativos se concreten- donde se concentra la mayor cantidad de incumplimientos.

### **8.2 SOBRE EL CUMPLIMIENTO DE LOS OBJETIVOS DEL PROYECTO**

El objetivo general de este trabajo consistió en *preparar fiscalizaciones preventivas efectivas y focalizadas en aquellos sostenedores y escuelas que tienen mayor probabilidad de estar utilizando indebidamente los recursos SEP; de esta manera, optimizar los recursos de la SUPEREDUC al momento de preparar el programa de fiscalización y la planificación de las visitas a terreno*. Para dar cumplimiento a este objetivo se definieron los siguientes objetivos específicos y cuyos resultados se presentan en lo sucesivo:

- *Segmentar y clasificar a sostenedores y escuelas en función de sus características y comportamientos relacionados con la gestión financiera, de modo que se tomen acciones preventivas para disminuir usos indebidos de recursos SEP a partir del*

*año 2018; utilizando para ello técnicas analíticas de data mining:* para ello, se rediseñó el procesos de fiscalización utilizando el patrón de negocio y arquetipo de procesos apropiado para emplear analítica en la preparación fiscalizaciones preventivas y focalizadas en aquellos sostenedores y/o escuelas que tienen mayor probabilidad de incumplir alguna materia en algún programa de fiscalización. En particular, este marco de trabajo al ser instanciado para el programa uso de recursos SEP logró segmentar a sostenedores y/o escuelas en función de sus características y comportamientos relacionados con la gestión financiera, identificando cinco clústeres de riesgo, de los cuales el clúster más riesgoso se compone de 1.962 escuelas que tras ser clasificadas, identificaron las 71 escuelas con la mayor probabilidad (cercana al 70%) de incumplir con el debido uso de recursos SEP; esta información, tras ser revisadas por un equipo de expertos de negocio se valida y se incluye en la muestra a fiscalizar del programa año 2018.

- *Generar recomendaciones oportunas a los sostenedores que se encuentran realizando la rendición de cuentas SEP, de modo de evitar y/o mitigar usos indebidos de recursos y alertar errores de imputación de gastos a partir del proceso declarativo año 2018; utilizando para ello técnicas analíticas de data mining:* para ello, se implementó un clasificador automático de textos basado en redes de palabras que fue piloteado con 41.171 descripciones de gastos de la cuenta de imputación combustible y peaje, y que tras validar sus resultados detectó más de 30.000 recomendaciones que de ser informadas oportunamente, hubiesen evitado incumplimientos en fondos SEP año 2017 por más de MM\$1.000. Los resultados obtenidos, la SUPEREDUC decidió incluirlos en la preparación de la muestra a fiscalizar para el presente año, y generar recomendaciones preventivas y/o de mitigación en esta cuenta de gasto a partir del proceso declarativo del año 2018 (que inicia en mes octubre) e incluir otras cuentas de gastos para el proceso declarativo del año 2019 (que iniciará junto con el año en mes de enero). Las recomendaciones, se incluyen como una línea de trabajo del plan *¡rinde bien, rinde seguro!* definido por la nueva administración.
- *Descubrir información en las descripciones de gastos de la rendición de cuentas SEP año 2017, para aumentar en al menos un 5% el monto en documentos*

*riesgosos respecto del año anterior y fiscalizarlos a partir de octubre año 2018:* para ello, se utiliza el clasificador de texto basado en redes de palabras y debido al arduo trabajo que conlleva pre procesar textos, preparar y clasificar los vectores de términos para cada tipo de recomendación (tipo 1: uso indebido de recursos SEP; tipo 2: error de imputación en cuenta de gasto; tipo 3: escasa descripción del gasto) y validar los resultados con expertos de negocio en cada nueva cuenta de gasto; se trabajó hasta cumplir el objetivo, el cual se logró al procesar las cuentas de gasto: transporte escolar, uniformes y vestuario, y combustible y peaje, y cuyos resultados serán incluidos en la preparación de la muestra a fiscalizar para el presente año . Las demás cuentas de gastos, serán trabajos futuros.

- *Aumentar la actual capacidad de visitas de fiscalización para el programa uso de recursos SEP en un 2% respecto del año 2017. Esta capacidad podrá ser utilizada en el primer año para reintegrar recursos SEP al sistema escolar, en especial aquéllos provenientes de documentos riesgosos detectados en punto anterior.* para ello, se implementó un *datamart* para trazar y dar visibilidad a los flujos de información desde que una visita a terreno, independiente del programa, es agendada hasta cerrar el servicio de fiscalización con un acta y calcular los tiempos promedios de duración del servicio. En particular, para el programa uso de recursos SEP tras ser monitoreado se concluye no estar optimizado, siendo posible aumentar la capacidad de fiscalización al menos un 2% al efectuar acciones de seguimiento y control al servicio; esto se corrobora al registrar los resultados en días que se efectuaron acciones de seguimiento y control en los cuales las duraciones fluctuaron de 2 a 3 días, mientras que en días sin seguimiento y control las duraciones fluctuaron de 4 a 10 días.

### **8.3 SOBRE LA CONTINUIDAD DE LAS TÉCNICAS DATA MINING EN LA SUPEREDUC**

La importancia de conocer a los clientes, segmentarlos y clasificarlos y ofrecerles servicios cada vez más personalizados no está suscrito sólo al ámbito privado; siendo, cada vez más, una obligación para los organismos públicos conocerlos y ofrecerles servicios de valor, dado que serán ellos quienes evaluarán los servicios prestados y se pronunciarán cada cuatro años (periodo de gobierno).



En el caso de las superintendencias, adicionalmente se deben conocer los sujetos de fiscalización, los cuales deben ser segmentados y clasificados según sus comportamientos; de forma tal, de focalizar los esfuerzos y recursos de la organización en aquellos que posean una mayor probabilidad de incumplimientos.

En este sentido, la SUPEREDUC mediante la metodología de extracción de conocimiento crisp-dm y las técnicas de segmentación y clasificación aplicada a sostenedores y/o escuelas en colaboración con fiscalizaciones preventivas y focalizadas, resulta ser una potente herramienta para el resguardo de derechos para los distintos actores de la comunidad educativa; junto con, optimizar los recursos de la Organización.

Cambiar de enfoque, en una institución que nace con un enfoque de fiscalización reactiva y basado en auditorías a uno cada vez más preventivo y centrado en los estudiantes, no es un proceso trivial y requiere de grandes esfuerzos de gestión del cambio y formación de capacidades al interior de la Organización que permitan dar continuidad a los trabajos realizados. En este sentido, la SUPEREDUC, comparada con otras superintendencias, ha invertido los recursos suficientes en hardware y software para aplicar analítica; junto con, capacitar a los miembros de la unidad de datos con el propósito de crear el ecosistema requerido para consolidar este tipo de herramientas. Además, la SUPEREDUC debe considerar seriamente utilizar este tipo de técnicas, ya que debe ser eficiente y eficaz en el resguardo de más de 1.800 obligaciones normativas y velar por el debido uso de más de 6 billones de pesos anuales o 3% del presupuesto del país, y todo con una capacidad limitada de visitas de fiscalización.

#### **8.4 SOBRE LAS TÉCNICAS UTILIZADAS PARA CONTRUIR EL CLASIFICADOR AUTOMÁTICO DE TEXTOS**

La metodología de extracción de conocimiento crisp-dm, las técnicas de text mining y redes de palabras permiten extraer información a partir de grandes volúmenes de textos en formato semi-estructurados (descripciones de gastos en las rendiciones de cuentas).

La etapa de pre procesamiento de textos es esencial para construir el diccionario semántico ad hoc a la problemática a resolver y al ámbito de negocio; junto con, aplicar correctores ortográficos, lista de términos stopword y n-grams que permiten reducir la

cardinalidad de términos. Con el propósito de aumentar la riqueza del diccionario, este se complementa en cada término con sus sinónimos, variantes y/o formas verbales.

Las redes de palabras construidas para cada categoría, a partir de vectores de términos pre procesados, limpios y clasificados por un experto, y sumado a la riqueza incorporada al diccionario, aumenta la probabilidad de encontrar vectores de términos semejantes no clasificados en la categoría; siendo este, un proceso iterativo.

Las redes de palabras al ser trabajadas como nodos y relaciones en una base de grafos (Neo4j) se facilita la aplicación de algoritmos de búsquedas (utilizando sentencias en Cypher) y posteriores procesamientos de resultados (utilizando API para java). De igual manera, tras identificar términos relevantes y/o discriminatorios a una categoría se aumenta considerablemente la efectividad y precisión de las búsquedas y su posterior clasificación en la categoría respectiva.

Respecto al funcionamiento del motor de búsqueda y clasificación, 40.000 textos estos son clasificados en aproximadamente 20 minutos usando un computador con un procesador core i7 con 8GB en RAM; lo cual, es aceptable al ser requerido un procesamiento en batch de las descripciones de gastos de las rendiciones de cuentas.

Finalmente, se concluye que es factible interpretar y clasificar una gran cantidad de documentos de textos a partir de las técnicas de text mining y redes de palabras, con una cantidad de recursos aceptable y en un tiempo considerablemente menor si esta tarea fuese trabajada manualmente.

## **8.5 SOBRE EL NUEVO SERVICIO DE RECOMENDACIONES A LAS RENDICIONES DE CUENTAS**

Este nuevo servicio al validar su potencial para reducir gradualmente año a año los actuales incumplimientos de sostenedores y/o escuelas respecto del uso de fondos SEP u otro tipo de recursos de fines específicos, la SUPEREDUC decide incorporarlo a los procesos de fiscalización de uso de recursos y asignar una persona como responsable.

En particular, este nuevo servicio aplicado oportunamente y durante el proceso declarativo de gastos de los recursos SEP permite disminuir el tipo de incumplimiento

*gasto no se ajusta al objeto de la subvención o aporte*, el cual representa más de un 60% de los incumplimientos en esta subvención y cuyos montos superan los MM\$ 9.817.

Este nuevo servicio, dota a la SUPEREDUC de la capacidad de generar durante el proceso de rendición de cuentas recomendaciones automatizadas oportunas a los sostenedores de modo de evitar y/o mitigar usos indebidos de recursos y alertar errores de imputación de gastos. Por ejemplo, para el programa de uso de recursos SEP, el clasificador automático de textos basado en redes de palabras, sólo en la cuenta de gasto combustible y peajes de más de 40.000 textos en año 2016, se pronuncia en más de 30.000 casos y donde en: 464 casos los clasifica del tipo 1: uso indebido de recursos SEP; 3.653 casos los clasifica del tipo 2: error de imputación en cuenta de gasto; 26.860 casos los clasifica del tipo 3: escasa descripción del gasto; cifras que al ser comparadas con año anterior, usando los métodos tradicionales, detectaron sólo 1.000 casos.

## **8.6 TRABAJOS FUTUROS**

- Completar el clasificador automático de textos basado en redes de palabras para las cuentas de gastos restantes para la rendición de fondos SEP, y lograr su máximo aprovechamiento a partir de la rendición de cuentas año 2019.
- Explorar nuevos criterios de riesgos que permitan detectar preventivamente aquellos sostenedores y/o documentos riesgosos vinculados a las siguientes categorías (ver Tabla N° 27) de incumplimientos: documentación enmendada o incorrecta, gasto rendido o pagado en período incorrecto, sin documentación de respaldo y documentación de respaldo incompleta; las que en su conjunto corresponden a más de un 30% de los incumplimientos en uso de fondos SEP.
- Con el conocimiento adquirido tras realizar este trabajo, y en particular para los programas uso de recursos, es factible diseñar e implementar un servicio de fiscalización digital que no requiera de visita a terreno y donde, por ejemplo, en un portal web privado para sostenedores, se publiquen ciertos tipos de documentos riesgosos y por cada uno, los sostenedores puedan aceptar el incumplimiento o adjuntar las evidencias y razones en su defensa.

## **CAPÍTULO 9: BIBLIOGRAFÍA**

Araya, J. (2012). Rediseño de Procesos de la Fiscalización Preventiva Dirección Nacional del Trabajo. Recuperado de <http://repositorio.uchile.cl/handle/2250/102788>

Banco Mundial (2016). Chile Panorama General. Recuperado de <http://www.bancomundial.org/es/country/chile/overview>

Barros, O. (2004): Ingeniería E-Business Ingeniería de Negocios para la Economía Digital. Santiago de Chile: J. C. Sáez Editor.

Barros, O. & Julio, C. (2011) Enterprise and Process Architecture Patterns", Business Process Management Journal. Recuperado de [https://www.researchgate.net/publication/235319849\\_Enterprise\\_and\\_Process\\_Architecture\\_Patterns](https://www.researchgate.net/publication/235319849_Enterprise_and_Process_Architecture_Patterns)

Barros, O. (2015). Ingeniería de Negocios: Diseño Integrado de Servicios, sus Procesos y Apoyo TI. Recuperado de [http://blog.obarros.cl/wp-content/uploads/2015/09/Ingenieriade-Negocios\\_OscarBarros2015MBE.pdf](http://blog.obarros.cl/wp-content/uploads/2015/09/Ingenieriade-Negocios_OscarBarros2015MBE.pdf)

Bittar, M. (2017). El Modelo PEST como Herramienta de Planificación. Recuperado de <http://www.elmundo.com.ve/firmas/moises-bittan/el-modelo-pest-como-herramienta-de-planificacion.aspx>

Cárdenas, J., Olivares, G., & Alfaro, R. (2014). Clasificación Automática de Textos Usando Redes de Palabras. Recuperado de <http://www.scielo.cl/pdf/signos/v47n86/a01.pdf>

Consejo para la Transparencia (2016). VII Estudio Nacional de Transparencia. Recuperado de [http://www.cooperativa.cl/noticias/site/artic/.../cp\\_estudio\\_nacional\\_transparencia.doc](http://www.cooperativa.cl/noticias/site/artic/.../cp_estudio_nacional_transparencia.doc)

Dirección de Presupuestos (2016). Informe de Presupuestos por Periodos. Recuperado de <http://www.dipres.gob.cl/595/w3-propertyname-540.html>

El Mostrador On Line (2014). Chile es un Líder Chile es líder regional en uso de Tecnologías de Información y Comunicación. Recuperado de <http://www.emol.com/noticias/tecnologia/2014/04/23/656703/chile-es-lider-regional-en-uso-de-tecnologias-de-informacion-y-comunicacion.html>

El Mercurio (2017). Recuperado de <http://impresa.elmercurio.com/pages/LUNHomepage.aspx?BodyID=2&dt=2017-09-25&dtB=2017-09-25>

Fahey, L. & Narayanan, V. (1986). *Macroenvironmental Analysis for Strategic Management*. USA: South-Western.

Farhangfara, A., Kurganb, L., & Dy, J. (2008). Impact of imputation of missing values on classification error for discrete data. *Pattern Recognition*, 41 (12), 3692-3705.

Fayyad, U. M., Piatetsky-Shapiro, G., & Smyth, P. (1996). Knowledge discovery and data mining: Towards a unifying framework. In *Knowledge Discovery and Data Mining*, 82–88.

Frakes, W. & Baeza-Yates, R. (1992). *Information Retrieval: Data Structures & Algorithms*. New Jersey: Prentice Hall.

Graham, R., Lenstra, j., & Tarjan, R. (1999). *Theory of Linear and Integer Programming*. Amsterdam: Wiley.

Han, J., Kamber, M. & Pei, J. (2012): *Data Mining Concepts and Techniques*. Waltham: Morgan Kaufmann Publishers.

Hotho, A. (2005). A Brief Survey of Text Mining. Recuperado de <https://http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.210.158&rep=rep1&type=pdf>

Jacobson, I., Booch, G. & Rumbaugh, J. (2000). *El Proceso Unificado de Desarrollo de Software*. Madrid: Addison Wesley.

Larman, C. (2010): *UML y Patrones Una Introducción al Análisis y Diseño Orientado a Objetos y al Proceso Unificado*. Madrid: Pearson Prentice Hall.

Ley N° 20.248, Subvención Escolar Preferencial. Publicado 25 enero de 2008. Recuperado de <https://www.leychile.cl/Navegar?idNorma=269001>

Ley N° 20.529, Sistema Nacional de Aseguramiento de la Calidad de la Educación Parvularia, Básica y Media y su Fiscalización. Publicado 27 de agosto de 2011. Recuperado de <http://www.leychile.cl/N?i=1028635&f=2012-12-11&p=>

Ley N° 20.730, Regula el lobby y las gestiones que representen intereses particulares ante autoridades y funcionarios. Publicado 8 de marzo de 2014. Recuperado de <https://www.leychile.cl/Navegar?idNorma=1060115>

Ley N° 20.845, De inclusión escolar que regula la admisión de los y las estudiantes, elimina el financiamiento compartido y prohíbe el lucro en establecimientos educacionales que reciben aportes del Estado. Publicado 8 de junio de 2015. Recuperado de <https://www.leychile.cl/Navegar?idNorma=1078172>

Lochbaum, K. & Streeter, L. (1989). Combining and comparing the effectiveness of latent semantic indexing and the ordinary vector space model for information retrieval. *Information Processing and Management*, 25(6):665–676.

Manual de Oslo, (2005). Recuperado de [www.madrid.org/bvirtual/BVCM001708.pdf](http://www.madrid.org/bvirtual/BVCM001708.pdf)

Ministerio de Economía, Fomento y Turismo (2013). Cuarta Encuesta Nacional sobre Gasto y Personal en Investigación y Desarrollo. Recuperado de <http://www.economia.gob.cl/estudios-y-encuestas/encuestas/encuestas-de-innovacion-e-id/encuesta-nacional-sobre-gasto-y-personal-en-investigacion-y-desarrollo/cuarta-encuesta-nacional-sobre-gasto-y-personal-en-investigacion-y-desarrollo-ano-de-referencia-2013>

Ministerio de Desarrollo Social (2015). Informe de Desarrollo Social 2015. Recuperado de <http://www.ministeriodesarrollosocial.gob.cl/pdf/upload/IDS2.pdf> Instituto Nacional de Estadísticas (2012). Resultados Preliminares Censo de Población y Vivienda 2012. Recuperado de [https://www.cooperativa.cl/noticias/site/artic/20120831/asocfile/20120831161553/resultados\\_preliminares\\_censo\\_2012.pdf](https://www.cooperativa.cl/noticias/site/artic/20120831/asocfile/20120831161553/resultados_preliminares_censo_2012.pdf)

Mitchell, T. (1997). *Machine Learning*. McGraw-Hill.

Norton, D. & Kaplan, R. (2011): *El Cuadro de Mando Integral The Balanced Scorecard*. Barcelona: Gestión 2000.

Osterwalder, A. & Pigneur, Y. (2009): *Generación de Modelos de Negocio*. Recuperado de <http://www.convergenciamultimedial.com/landau/documentos/bibliografia-2016/osterwalder.pdf>.

Pérez, V. (2011). Es necesario acelerar la inversión en Ciencia y Tecnología para Chile. Recuperado de <http://www.portaluchile.uchile.cl/portal/presentacion/rectoria/victor-perez-vera/discursos/discursos-academicos/64382/es-necesario-acelerar-la-inversion-en-ciencia-y-tecnologia-para-chile>

Programa de Naciones Unidas para el Desarrollo (2012). Informe sobre Desarrollo Humano en Chile 2012. Recuperado de <http://desarrollohumano.cl/idh/informes/2012-bienestar-subjetivo-el-desafio-de-repensar-el-desarrollo/>

Robertson, S. (1977). The probability ranking principle. *Journal of Documentation*, 33:294–304.

Ross, J., Weill, P & Robertson, D. (2006) *Enterprise Architecture as Strategy: Creating a Foundation for Business Execution*. Harvard Business School Press.

Salton, G., Wong, A., & Yang, C, S (1997). A vector space model for automatic indexing. *Communications of the ACM*, 18(11):613–620.

Superintendencia de Educación (2016). *La Institución*. Recuperado de <https://www.supereduc.cl/la-institucion/que-es-la-superintendencia/>

Ulrich, W. & Rosen, M. (2011). The Business Capability Map: The "Rosetta Stone" of Business/IT Alignment. Recuperado de <https://www.cutter.com/article/business-capability-map-rosetta-stone-businessit-alignment-469506>

Van Rijsbergen, C. (1986). A non-classical logic for information retrieval. *The Computer Journal*, 29(6):481–485.

## CAPÍTULO 10: ANEXOS

### ANEXO N° 1: ESTRUCTURA ORGÁNICA Y DOTACIÓN DE PERSONAL

En el sitio [http://transparencia.supereduc.cl/iris\\_gplus/Transparencia/indexa.jsp](http://transparencia.supereduc.cl/iris_gplus/Transparencia/indexa.jsp) se publica mensualmente toda la información asociada al personal de la SUPEREDUC; tal como ilustra la Figura N° 97.

The image shows a grid of information boxes on the SUPEREDUC Transparency Portal. The 'Estructura Orgánica' and 'Dotación de Personal' boxes are highlighted with a red border. The 'Estructura Orgánica' box contains links for 'Diagrama de la estructura orgánica' and 'Facultades, funciones y atribuciones de sus unidades u órganos internos'. The 'Dotación de Personal' box contains links for 'Dotación de Planta', 'Dotación a Contrata', 'Remuneraciones según estamentos, grados o jornadas', 'Dotación a Honorarios', and 'Dotación sujeta al Código del Trabajo'. Other visible boxes include 'Marco Normativo', 'Compras y Adquisiciones', 'Participación Ciudadana', 'Antecedentes preparatorios de normas', 'Información Presupuestaria', 'Subsidios y Beneficios', 'Declaración de patrimonio e intereses', 'Auditorías al Ejercicio Presupuestario', 'Ley del Lobby', 'Vínculos Institucionales', 'Derecho de Acceso a la Información Pública', 'Trámites del Organismo', and 'Transferencias'.

Figura N° 97: Sitio de Gobierno Transparente de la SUPEREDUC



El Organigrama Institucional, es ilustrado en la Figura N° 98.

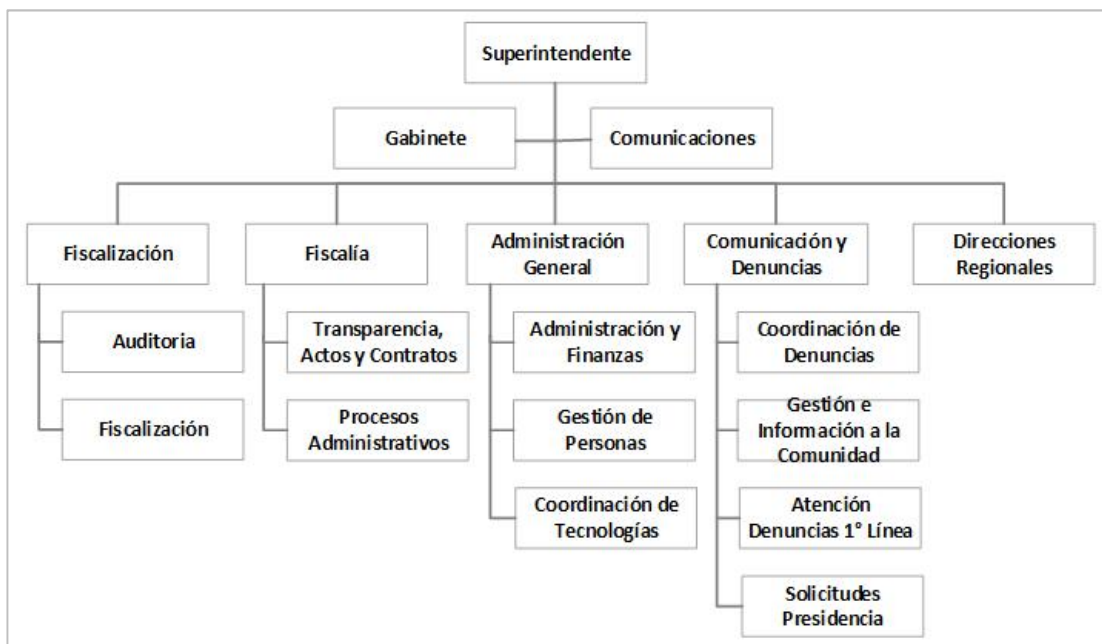


Figura N° 98: Organigrama de la SUPEREDUC

## ANEXO N° 2: BALANCED SCORECARD DE LA SUPEREDUC

Tabla N° 30: Valor Político

N°	Objetivos Estratégicos	Indicadores	Metas
1	Mejorar el uso de los recursos asignados para fines educativos	% de recursos públicos transferidos declarados en el sistema de rendición de cuentas en el año.	>=95%
		% de recursos del sistema rendidos y gastados según normativa pertinente del total de EE fiscalizados.	Por definir
		% de asignaciones presupuestarios de subvenciones que se someten al proceso de rendición de cuentas del año t, de recursos percibidos por los sostenedores de establecimientos en el año t-1.	100%
2	Posicionar a la Superintendencia como una institución moderna, proba y profesional	% de recursos que presenten sus rendiciones basadas en Estados Financieros.	Por definir
		% de mejora respecto al estudio de percepción y conocimiento de la institución.	Por definir
		Protocolos de probidad actualizados y difundidos hacia la comunidad educativa y comunidad en general.	Una vez al año
		Carta de derechos y deberes ciudadanos publicada.	Sí
		% de solicitudes de acceso a la información pública (Ley 20.285) respondidas en plazo legal.	100%
		% de resoluciones del Consejo de Transparencia del Estado adversas a la Superintendencia.	Por definir
3	Avanzar eficazmente en la puesta en marcha del nuevo marco normativo de la reforma	% de recursos ratificados por la corte de apelaciones.	Por definir
		% de observaciones del informe final de CGR subsanadas.	100%
		% de implementación del plan referido a LIE.	100%
4	Mejorar el cumplimiento normativo de los establecimientos, priorizando los aspectos y criterios que aseguren la calidad en la educación	% de implementación del plan referido a Parvularia.	100%
		% de implementación del plan de difusión del nuevo marco normativo.	100%
		% de implementación del Nuevo Modelo de Fiscalización.	Por definir
5	Consolidar a la Superintendencia como agente activo del SAC	% de mejora en el cumplimiento normativo	Por definir
		% cumplimiento de los compromisos de la Superintendencia en el Plan SAC.	100%
		% implementación de campaña de difusión interna sobre rol de la Superintendencia en el SAC.	100%

Tabla N° 31: Valor Beneficiario

N°	Objetivos Estratégicos	Indicadores	Metas
6	Responder oportuna y eficazmente los reclamos y denuncias de los ciudadanos	% cumplimiento tiempo de respuesta tipificado	Por definir
		% de avance de implementación del "Árbol Integral de Atención Ciudadana con plazos y roles responsables"	100%
		% de mejora del tiempo promedio del los casos tipificados que están en cada división	Por definir
7	Mejorar las capacidades y prácticas de las escuelas en el ámbito de las competencias de la Superintendencia	Catálogo de buenas practicas y capacidades de establecimientos en el ámbito de las competencias de la Superintendencia publicado	Sí
		% de ejecución del plan de desarrollo de capacidades para los establecimientos	100%
		Modelo de relación con la comunidad escolar implementado	Sí
8	Aumentar el conocimiento de los derechos educativos dentro de la comunidad escolar	% de mejora del nivel de conocimiento de los derechos educativos dentro de la comunidad escolar	Por definir
9	Proveer información a los actores del SAC que agregue valor	Área de estudio y estadísticas implementada	Sí
		Set de indicadores de cumplimiento normativo y rendición de cuentas definidos	Sí

Tabla N° 32: Valor Interno

N°	Objetivos Estratégicos	Indicadores	Metas
10	Contar con procesos integrados, efectivos y con foco en la probidad	Cadena de Valor Definida	Sí
		% Diseñar e implementar Gestión por Procesos	100%
		Instancia de interoperatividad implementada	Sí
		% implementación del Arbol de atención ciudadana integrado a sustentos de Fiscalización y Graduador de Fiscalía	100%
11	Generar una organización abierta a la gestión del cambio, con regiones activas, que cuente con la dotación y estructura adecuada y sea capaz de gestionar el conocimiento y el aprendizaje	Documento que resume buenas prácticas al interior de cada División y Dirección Regional difundido	Sí
		Proceso de planificación de fiscalización con variables territoriales incorporadas	Sí
		% de implementación de estructura organizacional	100%
		Metodología de gestión del cambio definida	Sí
12	Generar un modelo de relación externa con los organismos relacionados a la Superintendencia	N° de convenios firmados con otros servicios	Por definir
		Modelo de relación con actores externos implementado	Sí

Tabla N° 33: Valor Futuro

N°	Objetivos Estratégicos	Indicadores	Metas
13	Contar con funcionarios capacitados y motivados para contribuir a la mejora continua	% de ejecución del Plan Anual de Capacitación	95%
		% funcionarios capacitados del total programado	90%
		% de Ejecución Presupuestaria del PAC	97%
		% de actividades ejecutadas del PAC, con encuestas de satisfacción realizada	95%
		% de funcionarios que participan en alguna jornada de inducción durante el año	90%
		N° de jornada de inducción en el año	3
		% de aumento de satisfacción del plan de capacitaciones	Por definir
14	Desarrollar una cultura interna basada en la colaboración, comunicación, el bienestar y sentido de pertenencia	PAC alineado con los nuevos desafíos y carrera profesional para el 2016	Sí
		Política de bienestar de la institución definida y difundida	Sí
		% de mejora en la satisfacción de canales y contenidos de información institucional	Por definir
		% de plan de desarrollo de cultura organizacional ejecutado	100%
		Encuesta de clima laboral ejecutada	>=2
		% de mejora en la encuesta de clima laboral	Por definir
		Unidad de Relaciones Laborales implementada	Sí
15	Potenciar las habilidades directivas que soporten la gestión del cambio	% de implementación de buenas prácticas en la institución	100%
		Programa de Habilidades Directivas diseñado y validado por el Superintendente	Sí
		% de implementación del Programa de Habilidades Directivas	100%
		% de implementación del plan de cohesión del equipo directivo	100%

### ANEXO N° 3: CONCENTRACIÓN DE VALORES EN VARIABLES

- Variable: Dependencia

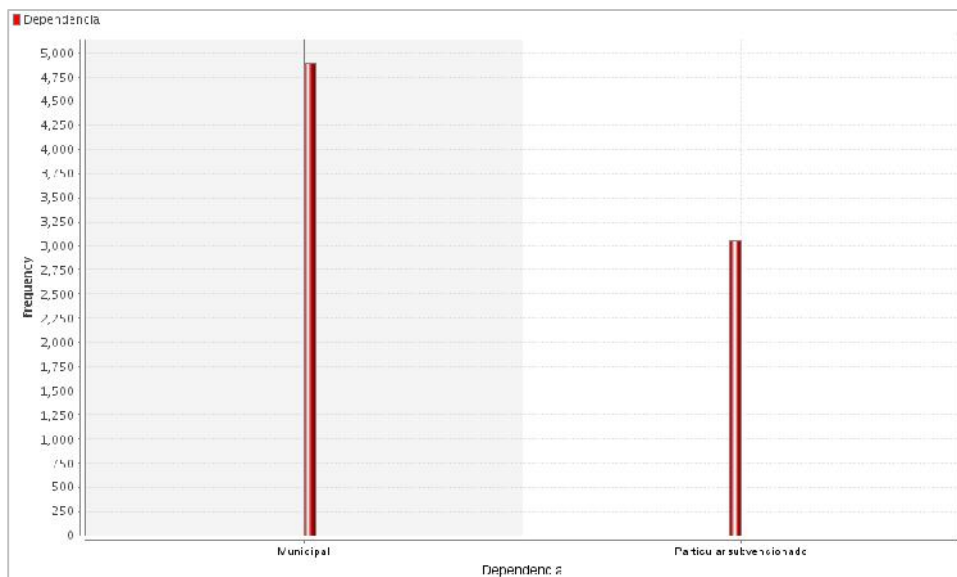


Figura N° 99: Dependencia administrativa escuelas SEP

- Variable: Zona

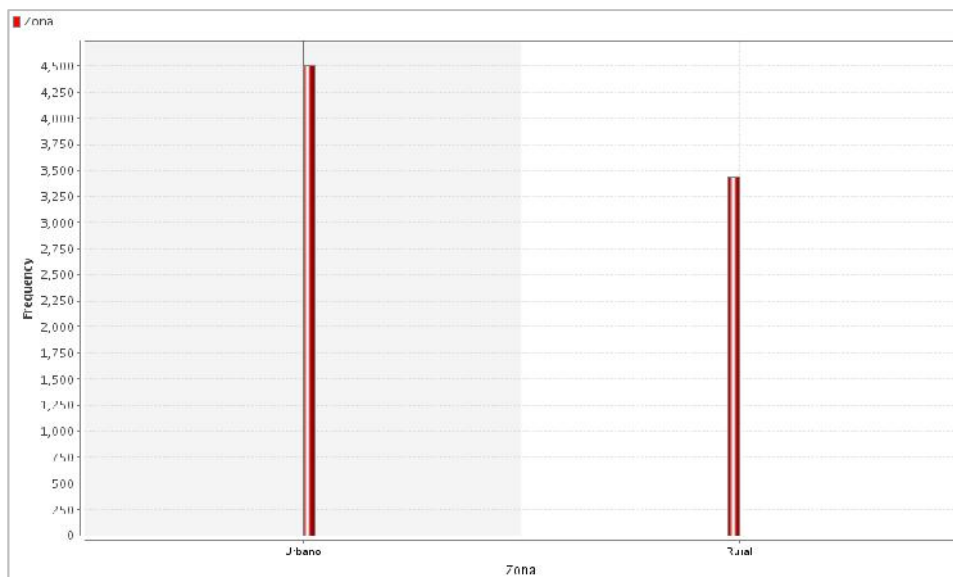


Figura N° 100: Zona geográfica escuelas SEP

- Variable: Clasificación

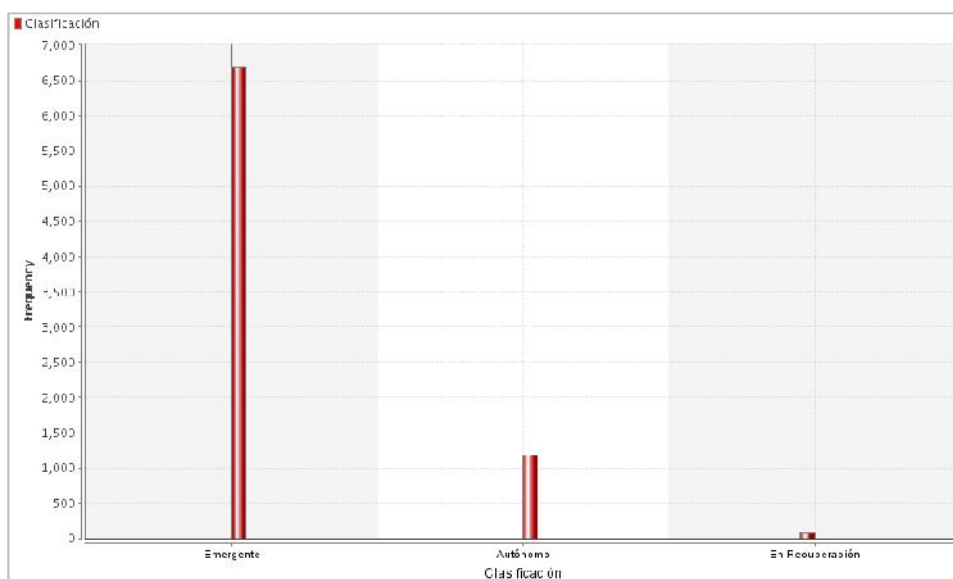


Figura N° 101: Clasificación<sup>30</sup> de escuelas SEP

- Variable: Parvularia

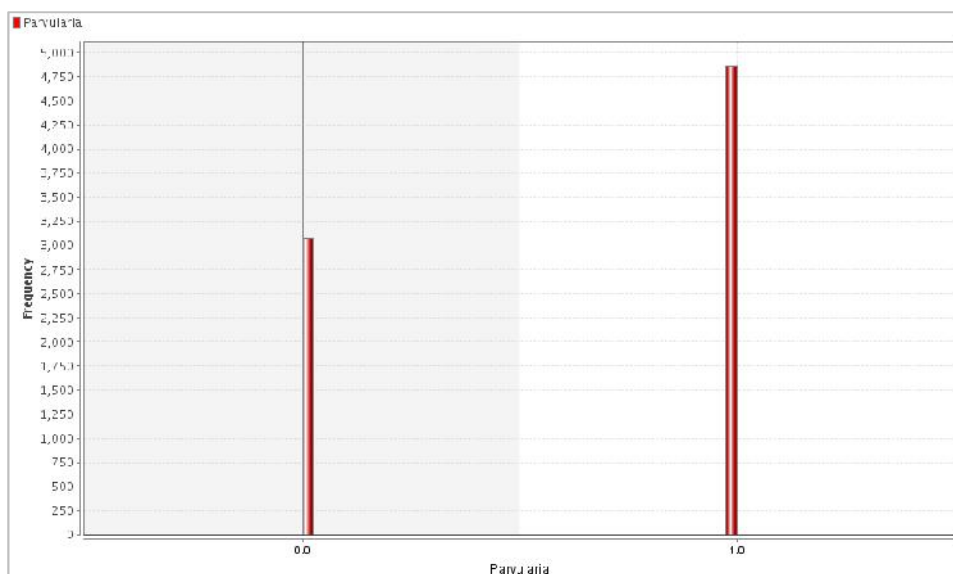


Figura N° 102: Educación parvularia en escuelas SEP

---

<sup>30</sup> En el sitio <https://www.ayudameduc.cl/ficha/clasificacion-de-establecimientos-sep-10> se encuentra información de las clasificación de escuelas y como se realiza el proceso.

- Variable: Básica

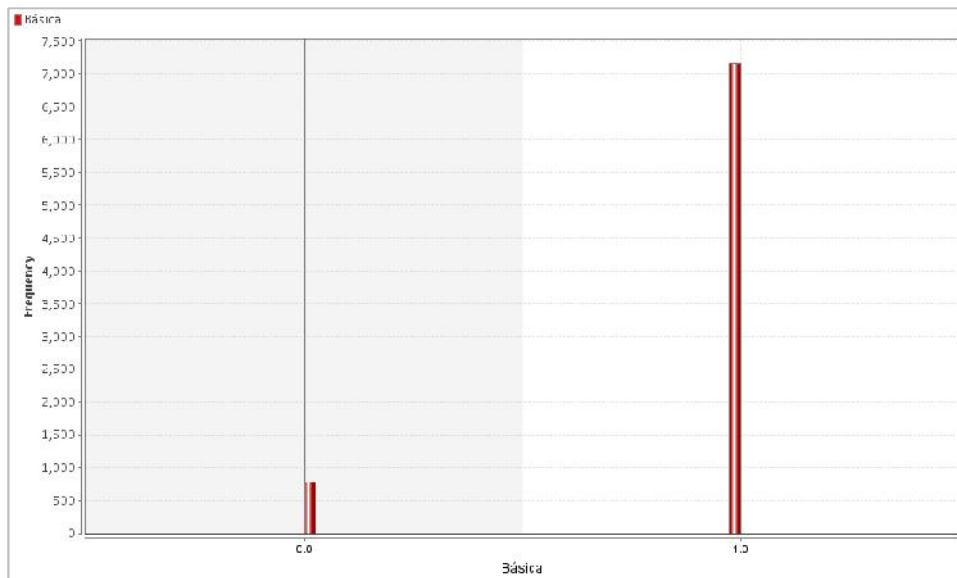


Figura N° 103: Educación básica en escuelas SEP

- Variable: Media HC

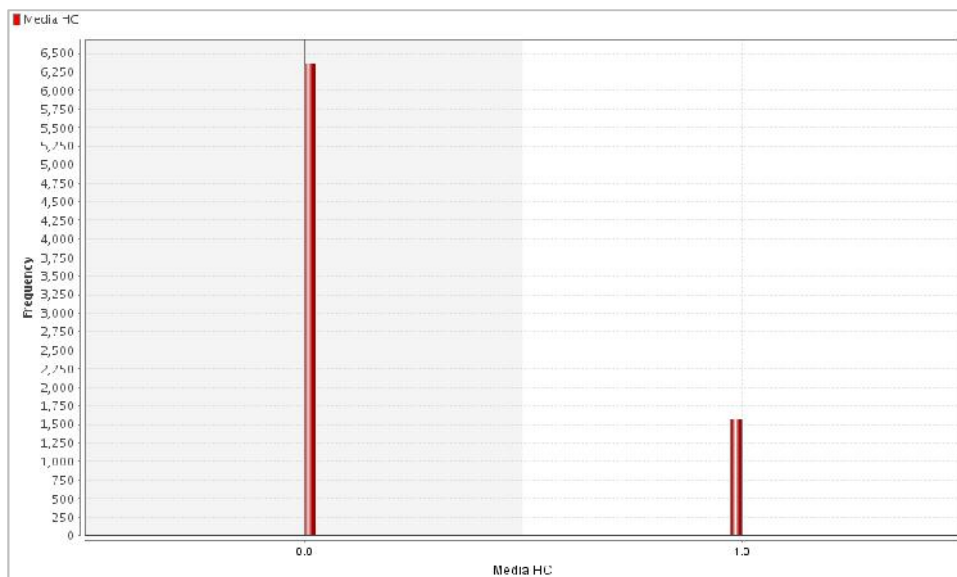


Figura N° 104: Educación media científico humanista escuelas SEP

- Variable: Media TP

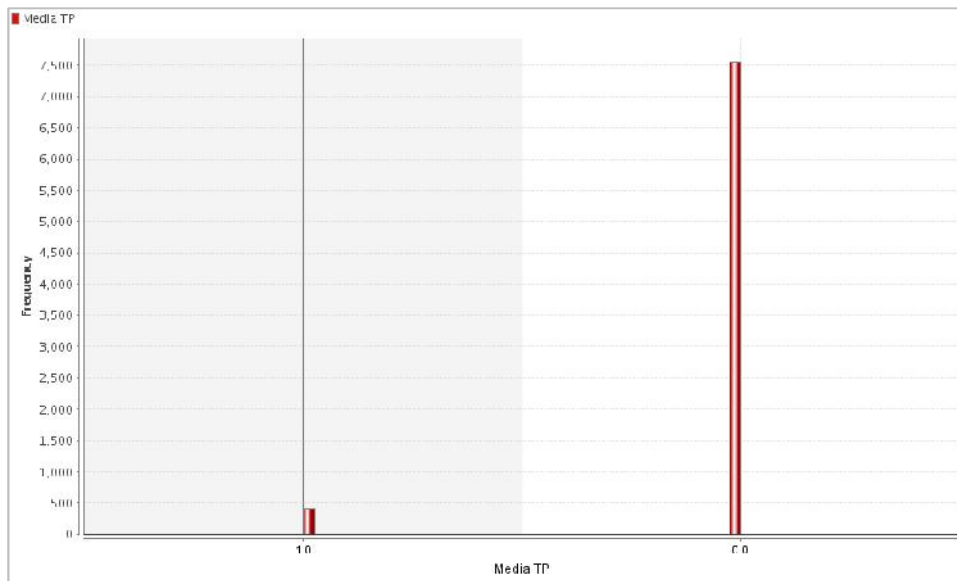


Figura N° 105: Educación media técnico profesional escuelas SEP

- Variable: Matrícula

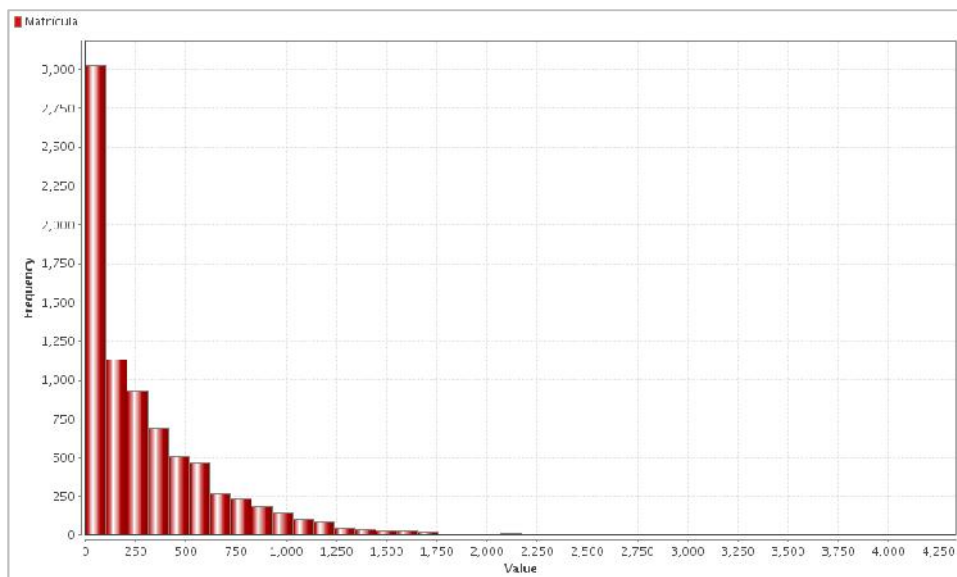


Figura N° 106: Matrícula año 2016 escuelas SEP

- Variable: Cursos

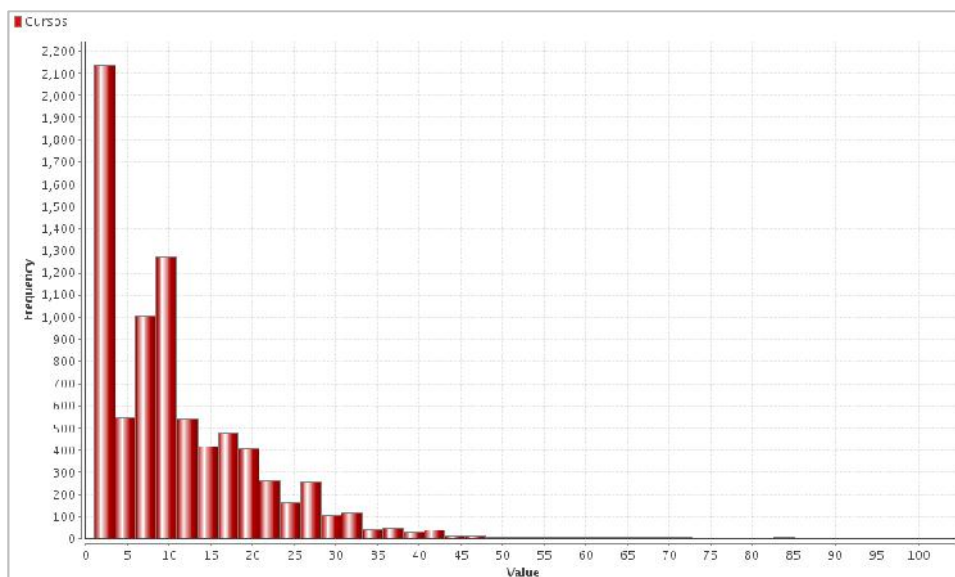


Figura N° 107: Cantidad de cursos año 2016 escuelas SEP

- Variable: Matrícula – Curso

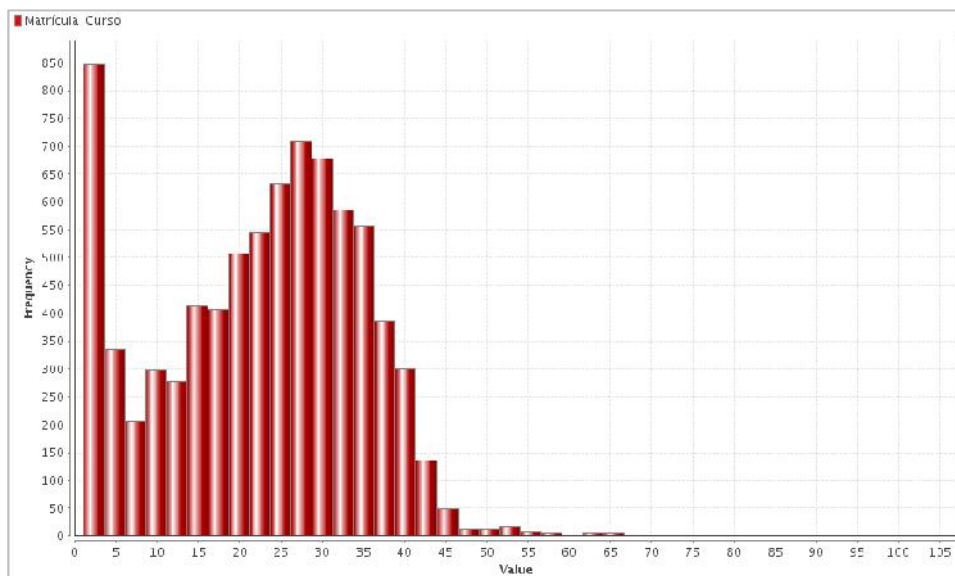


Figura N° 108: Matrícula promedio por curso escuelas SEP



- Variable: Ganancia Matrícula

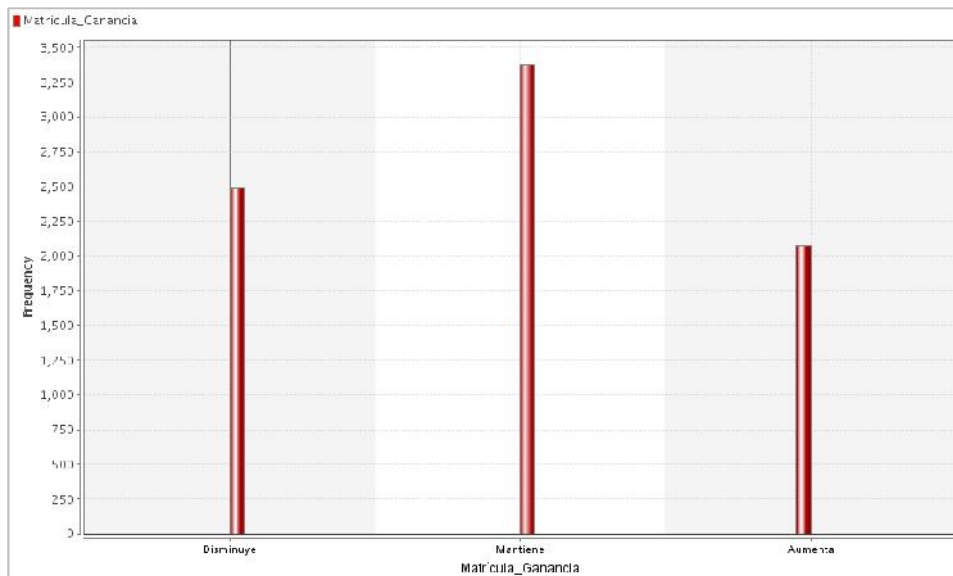


Figura N° 109: Ganancia de matrícula escuelas SEP, año 2016 respecto 2015

- Variable: Docentes

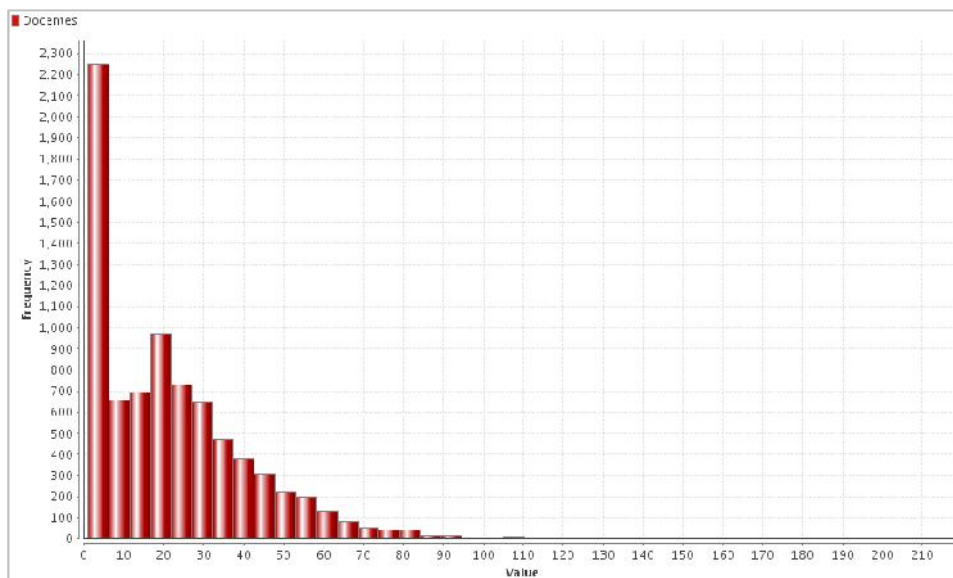


Figura N° 110: Cantidad de docentes en escuelas SEP

- Variable: Asistentes

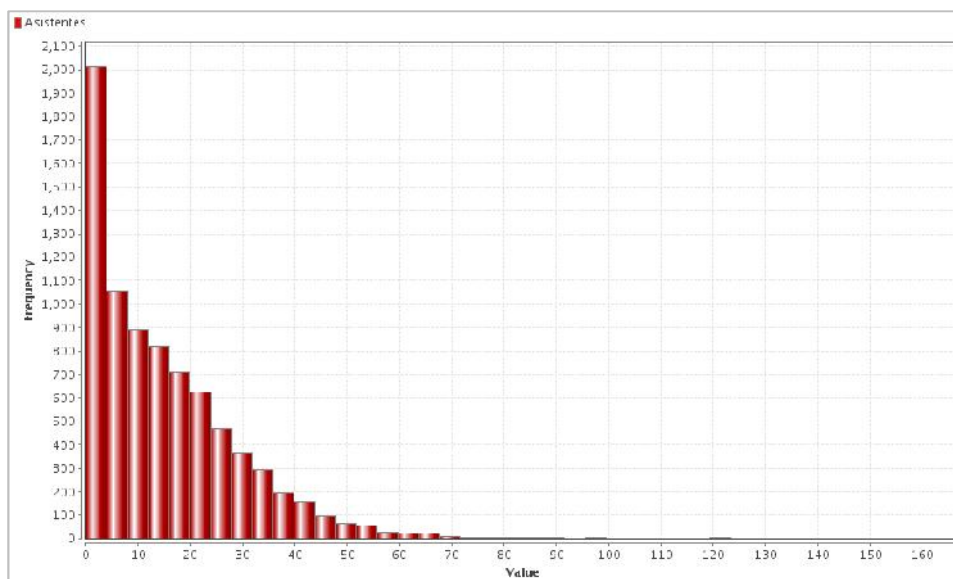


Figura N° 111: Cantidad de asistentes de la educación en escuelas SEP

- Variable: Planta

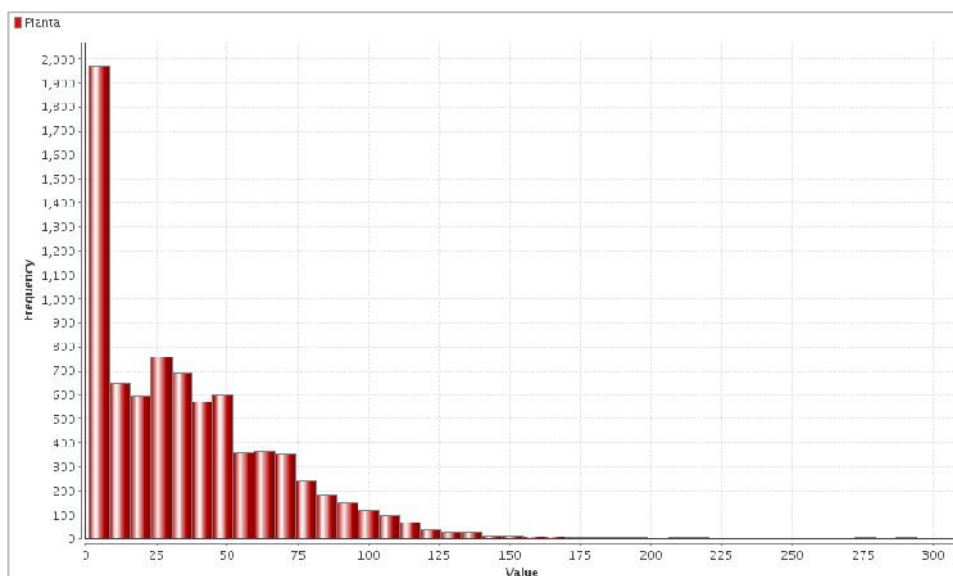


Figura N° 112: Cantidad de docentes y asistentes de la educación en escuelas SEP

- Variable: Ganancia Planta

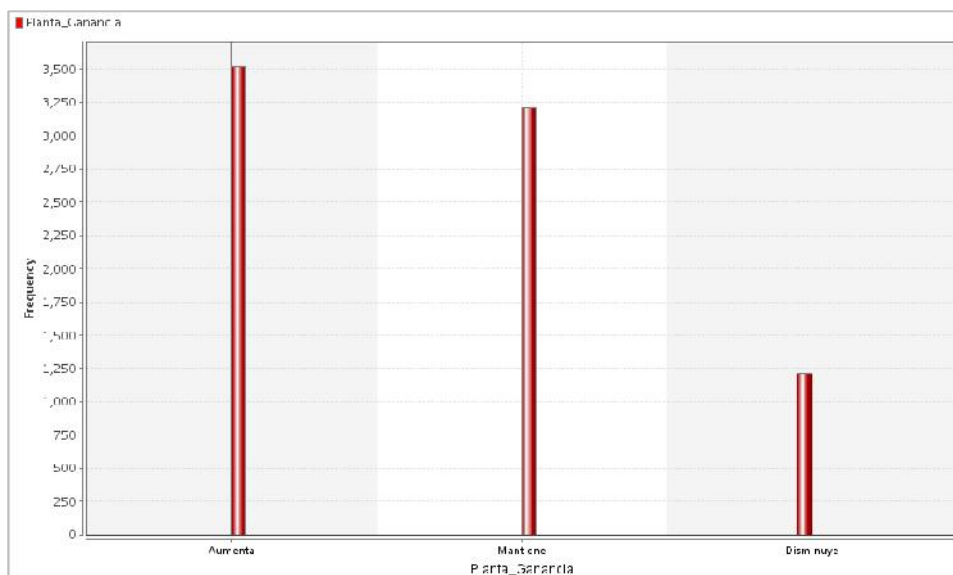


Figura N° 113: Ganancia de planta escuelas SEP, año 2016 respecto año 2015

- Variable: Gestión Financiera

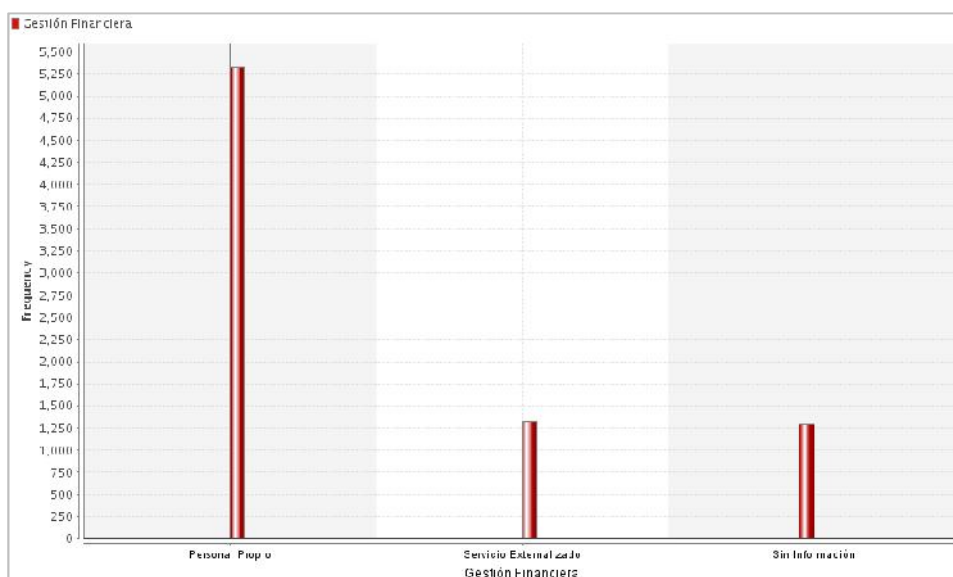


Figura N° 114: Tipo de apoyo en gestión financiera escuelas SEP

- Variable: Gasto Planta

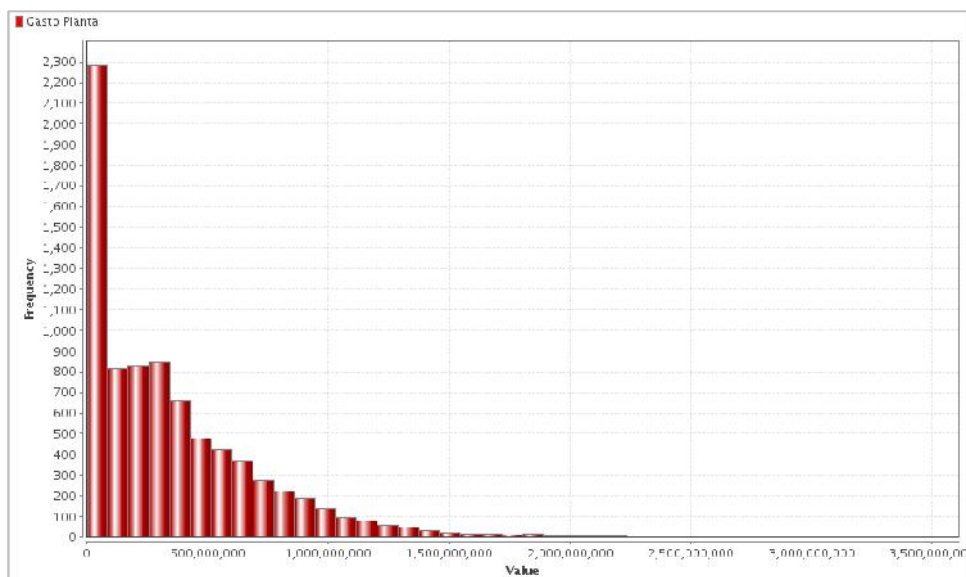


Figura N° 115: Gasto incurrido en planta año 2016 en escuelas SEP

- Variable: Ingresos

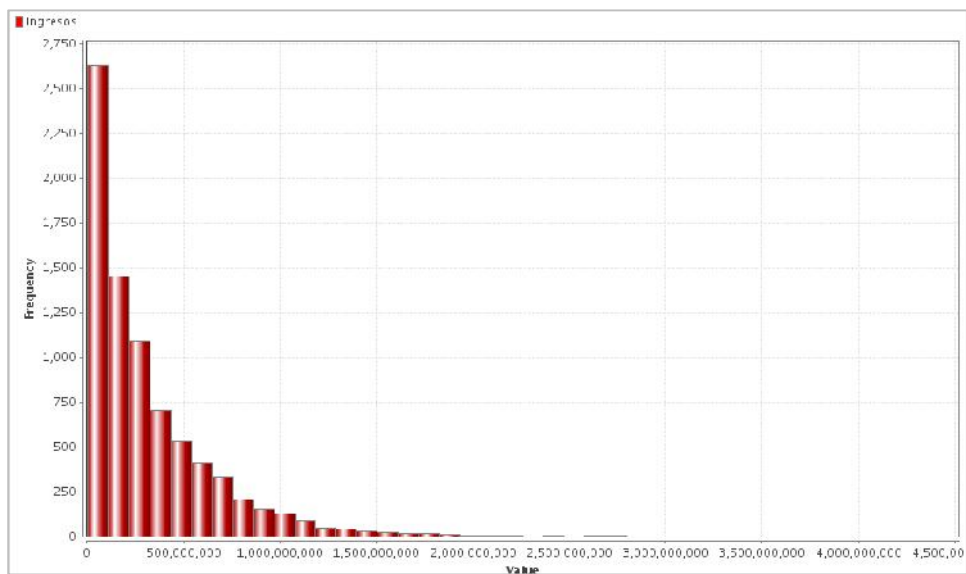


Figura N° 116: Ingresos subvención normal año 2016 escuelas SEP

- Variable: Planta – Ingresos

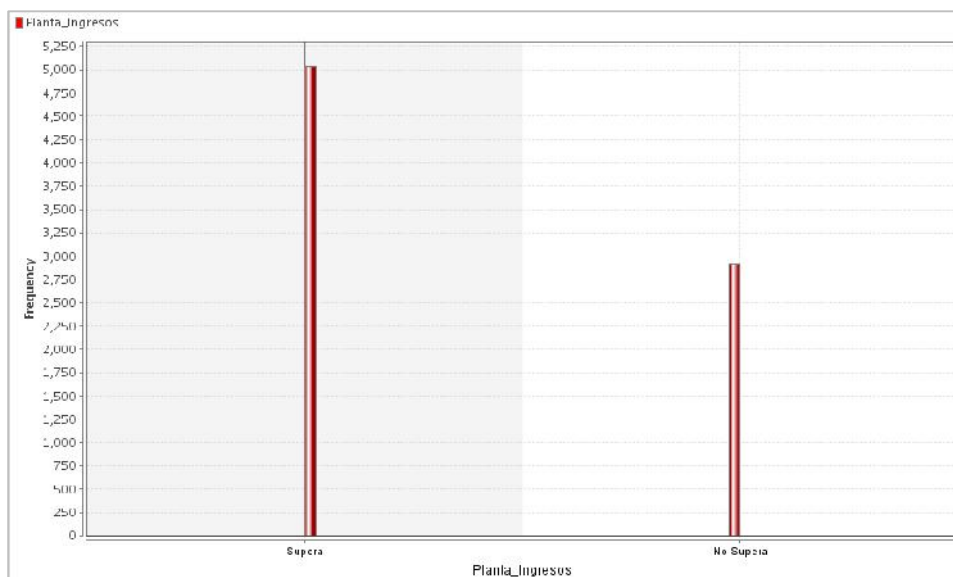


Figura N° 117: Relación costo de la planta versus los ingresos en escuelas SEP

- Variable: Gastos Rechazados año 2014

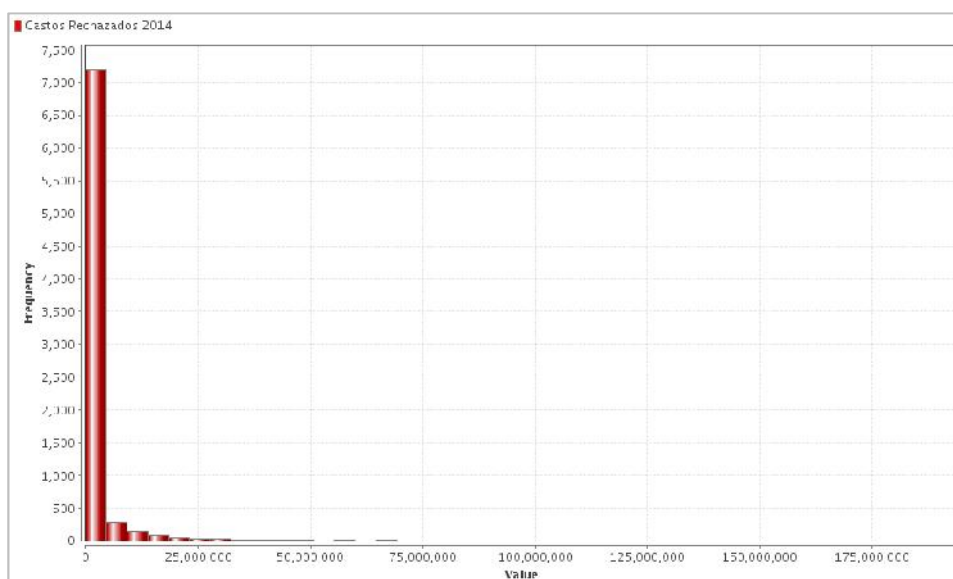


Figura N° 118: Gastos rechazados año 2014 en escuelas SEP

- Variable: Gastos Rechazados año 2015



Figura N° 119: Gastos rechazados año 2015 en escuelas SEP

- Variable: Multas Año 2014

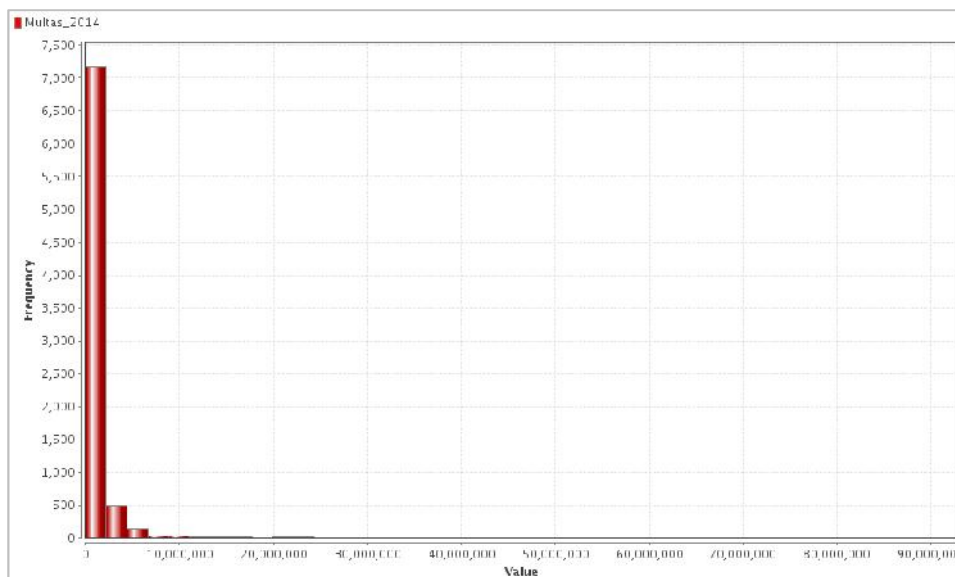


Figura N° 120: Multas año 2014 escuelas SEP

- Variable: Multas Año 2015

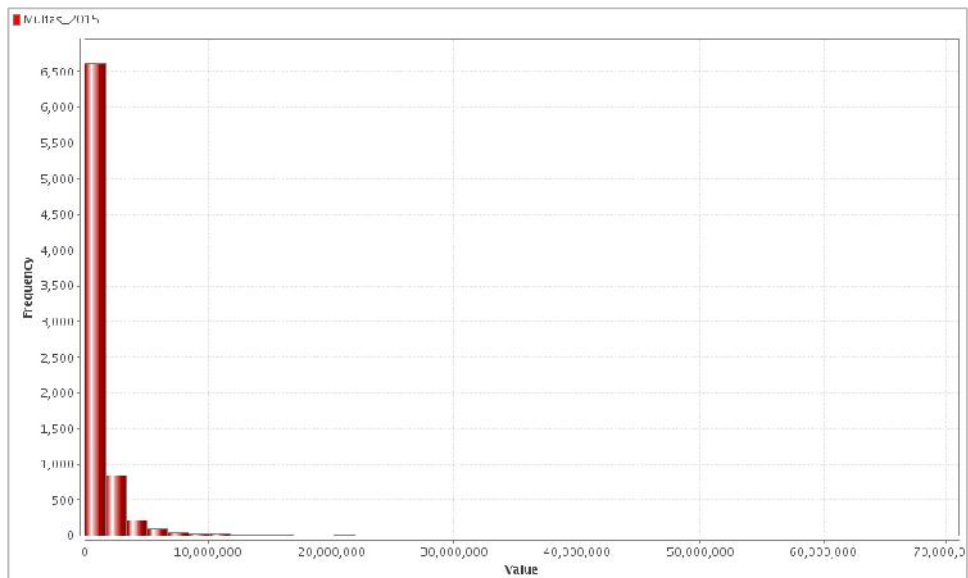


Figura N° 121: Multas año 2015 escuelas SEP

- Variable: Multas Año 2016

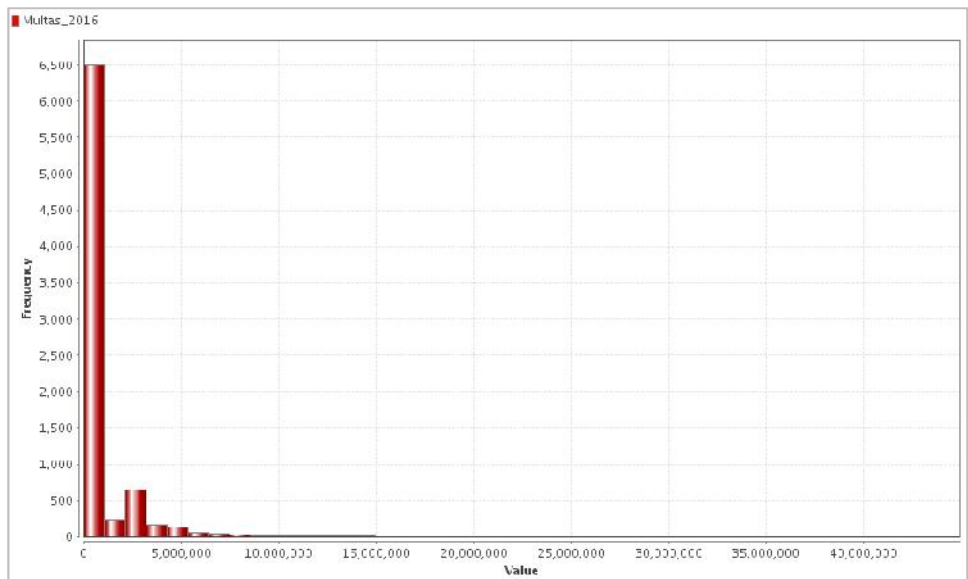


Figura N° 122: Multas año 2016 escuelas SEP

- Variable: Retenciones Año 2014

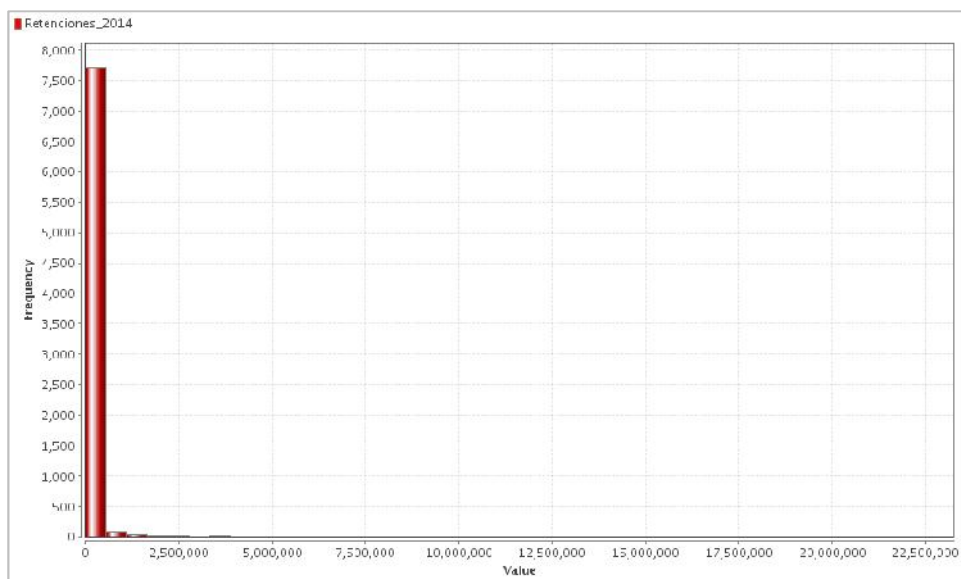


Figura N° 123: Retenciones año 2014 en escuelas SEP

- Variable: Retenciones Año 2015

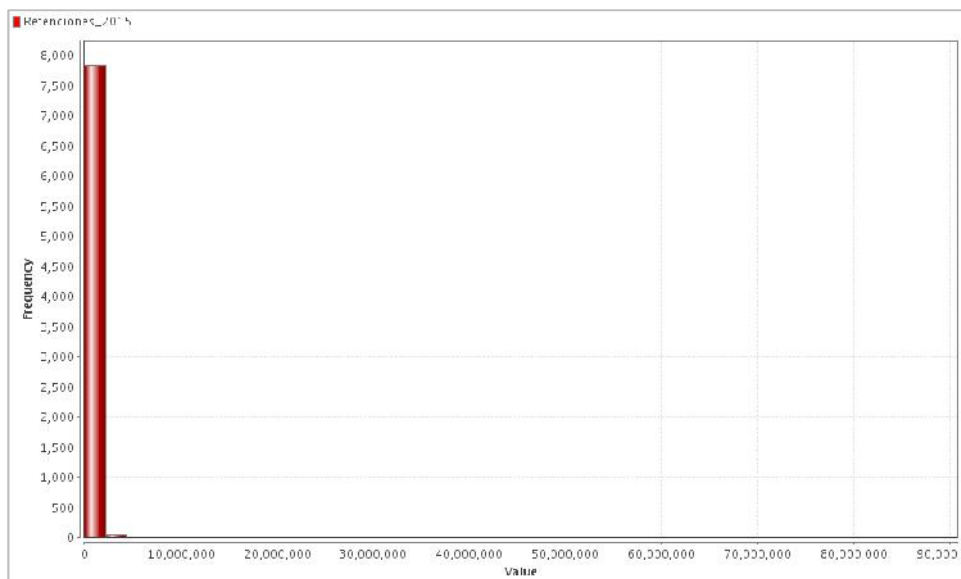


Figura N° 124: Retenciones año 2015 en escuelas SEP



- Variable: Retenciones Año 2016



Figura N° 125: Retenciones año 2016 en escuelas SEP

## ANEXO N° 4: DETALLES ALGORITMO DE OPTIMIZACIÓN

### Sets

j escuelas /esc1\*esc20 / ;

### Parameters

b(j) reintegro MM\$ de la escuela j

/ esc1	73.8
esc2	88.6
esc3	91.8
esc4	29.7
esc5	78
esc6	96.4
esc7	109.6
esc8	123.2
esc9	102.6
esc10	103.1
esc11	99.9
esc12	69.3
esc13	61.8
esc14	73.7
esc15	50.2
esc16	34.3
esc17	114.2
esc18	98.1
esc19	95.9
esc20	93.3 /

n(j) nunca fiscalizadas

/	esc1	0	
	esc2	1	
	esc3	0	
	esc4	0	
	esc5	0	
	esc6	1	
	esc7	0	
	esc8	1	
	esc9	0	
	esc10	0	
	esc11	1	
	esc12	0	
	esc13	0	
	esc14	0	
	esc15	0	
	esc16	1	
	esc17	0	
	esc18	0	
	esc19	0	
	esc20	1	/

ps(j) particulare subvencionadas

/	esc1	0
	esc2	0
	esc3	1
	esc4	0
	esc5	1
	esc6	1

esc7	1
esc8	1
esc9	1
esc10	0
esc11	1
esc12	0
esc13	1
esc14	0
esc15	1
esc16	1
esc17	1
esc18	0
esc19	1
esc20	1 /

m(j) municipales

/ esc1	1
esc2	1
esc3	0
esc4	1
esc5	0
esc6	0
esc7	0
esc8	0
esc9	0
esc10	1
esc11	0
esc12	1
esc13	0

esc14 1  
 esc15 0  
 esc16 0  
 esc17 0  
 esc18 1  
 esc19 0  
 esc20 0 /

f(j) fijadas

/ esc3 1  
 esc17 1  
 esc20 1 / ;

Scalar CTVR cantidad de visitas asignadas a la zona /10/ ;

Scalar CENFR cantidad de escuelas nunca fiscalizadas a incluir en la zona /3/;

Scalar CEPSSR cantidad de escuelas particular subvencionada a incluir en la zona /6/;

Scalar CEMR cantidad de escuelas municipales a incluir en la zona /4/;

Scalar FIXA cantidad de escuelas fijada a incluir en la zona /3/;

Variables

e(j) escuelas de la zona incluida en la muestra

z reintegro total MM\$ ;

Binary Variables e(j) ;

Equations

reintegro define la función objetivo

visitas escuelas muestra

fija escuelas fijadas

nunca      escuelas nunca fiscalizadas  
particular    escuelas particulares subvencionadas  
municipal    escuelas municipales ;

reintegro ..       $z = e = \text{sum}((j), b(j)*e(j)) ;$

visitas ..       $\text{sum}((j), e(j)) = e = \text{CTVR} ;$

fija ..       $\text{sum}((j), f(j)*e(j)) = e = \text{FIXA} ;$

nunca ..       $\text{sum}((j), n(j)*e(j)) = l = \text{CENFR} ;$

particular ..     $\text{sum}((j), ps(j)*e(j)) = l = \text{CEPSR} ;$

municipal ..     $\text{sum}((j), m(j)*e(j)) = l = \text{CEMR} ;$

Model muestra /all/ ;

Solve muestra using MIP maximizing z ;